(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 September 2018 (27.09.2018)

W!P O | PCT

(10) International Publication Number
WO 2018/176009 Al

(51) International Patent Classification:
C12N 15/11 (2006.01) *C12N 15/10* (2006.01)

(21) International Application Number:
PCT/US20 18/024208

(22) International Filing Date:
23 March 2018 (23.03.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/475,830	23 March 2017 (23.03.2017)	US
62/490,587	26 April 2017 (26.04.2017)	US
62/5 11,934	26 May 2017 (26.05.2017)	**us**
62/55 1,95 1	30 August 2017 (30.08.2017)	**us**

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **LIU, David, R.;** 3 Whitman Circle, Lexington, MA 02420 (US). **KOMOR, Alexis, Christine;** 9135 Judicial Drive, #3419, San Diego, CA 92122 (US). **CHEN, Liwei;** 86 Avon Street, Apt. 2, Somerville, MA 02143 (US). **REES, Holly, A.;** 34 Hawthorne Street, Somerville, MA 02144 (US).

(74) Agent: **HEBERT, Alan, M.;** Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(54) Title: NUCLEOBASE EDITORS COMPRISING NUCLEIC ACID PROGRAMMABLE DNA BINDING PROTEINS

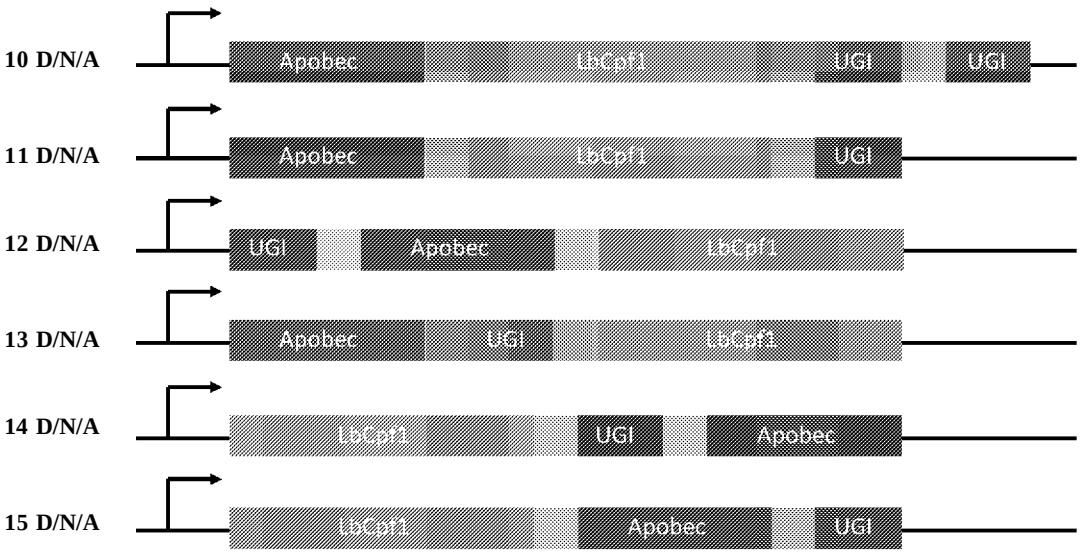


FIGURE 152

(57) Abstract: Some aspects of this disclosure provide strategies, systems, reagents, methods, and kits that are useful for the targeted editing of nucleic acids, including editing a single site within the genome of a cell or subject, e. g., within the human genome. In some embodiments, fusion proteins of nucleic acid programmable DNA binding proteins (napDNAbp), e. g., Cpf 1 or variants thereof, and nucleic acid editing proteins or protein domains, e. g., deaminase domains, are provided. In some embodiments, methods for targeted nucleic acid editing are provided. In some embodiments, reagents and kits for the generation of targeted nucleic acid editing proteins, e. g., fusion proteins of a napDNAbp (e. g., CasX, CasY, Cpfl, C2cl, C2c2, C2C3, and Argonaute) and nucleic acid editing proteins or domains, are provided.

WO 2018/176009 A1

[Continued on next page]

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

NUCLEOBASE EDITORS COMPRISING NUCLEIC ACID PROGRAMMABLE DNA BINDING PROTEINS

BACKGROUND OF THE INVENTION

[0001] Targeted editing of nucleic acid sequences, for example, the targeted cleavage or the targeted introduction of a specific modification into genomic DNA, is a highly promising approach for the study of gene function and also has the potential to provide new therapies for human genetic diseases.[1] An ideal nucleic acid editing technology possesses three characteristics: (1) high efficiency of installing the desired modification; (2) minimal off-target activity; and (3) the ability to be programmed to edit precisely any site in a given nucleic acid, *e.g.,* any site within the human genome.[2] Current genome engineering tools, including engineered zinc finger nucleases (ZFNs),[3] transcription activator like effector nucleases (TALENs),[4] and most recently, the RNA-guided DNA endonuclease Cas9,[5] effect sequence-specific DNA cleavage in a genome. This programmable cleavage can result in mutation of the DNA at the cleavage site via non-homologous end joining (NHEJ) or replacement of the DNA surrounding the cleavage site via homology-directed repair (HDR).[6,7]

[0002] One drawback to the current technologies is that both NHEJ and HDR are stochastic processes that typically result in modest gene editing efficiencies as well as unwanted gene alterations that can compete with the desired alteration.[8] Since many genetic diseases in principle can be treated by effecting a specific nucleotide change at a specific location in the genome (for example, a C to T change in a specific codon of a gene associated with a disease),[9] the development of a programmable way to achieve such precision gene editing would represent both a powerful new research tool, as well as a potential new approach to gene editing-based human therapeutics.

SUMMARY OF THE INVENTION

[0003] Nucleic acid programmable DNA binding proteins (napDNAbp), such as the clustered regularly interspaced short palindromic repeat (CRISPR) system is a recently discovered prokaryotic adaptive immune system[10] that has been modified to enable robust and general genome engineering in a variety of organisms and cell lines.[11] CRISPR-Cas (CRISPR associated) systems are protein-RNA complexes that use an RNA molecule (sgRNA) as a guide to localize the complex to a target DNA sequence *via* base-pairing.[12] In the natural systems, a Cas protein then acts as an endonuclease to cleave the targeted DNA sequence.[13] The target DNA sequence must be both complementary to the sgRNA, and also contain a

"protospacer-adjacent motif (PAM) at the 3'-end of the complementary region in order for the system to function.[14]

[0004] Among the known Cas proteins, *S. pyogenes* Cas9 has been mostly widely used as a tool for genome engineering.[15] This Cas9 protein is a large, multi-domain protein containing two distinct nuclease domains. Point mutations can be introduced into Cas9 to abolish nuclease activity, resulting in a dead Cas9 (dCas9) that still retains its ability to bind DNA in a sgRNA-programmed manner.[16] In principle, when fused to another protein or domain, dCas9 can target that protein or domain to virtually any DNA sequence simply by co-expression with an appropriate sgRNA.

[0005] The potential of the dCas9 complex for genome engineering purposes is immense. Its unique ability to bring proteins to specific sites in a genome programmed by the sgRNA in theory can be developed into a variety of site-specific genome engineering tools beyond nucleases, including deaminases (e.g., cytidine deamianses), transcriptional activators, transcriptional repressors, histone-modifying proteins, integrases, and recombinases.[11] Some of these potential applications have recently been implemented through dCas9 fusions with transcriptional activators to afford RNA-guided transcriptional activators,[17,18] transcriptional repressors,[16,19,20] and chromatin modification enzymes.[21] Simple co-expression of these fusions with a variety of sgRNAs results in specific expression of the target genes. These seminal studies have paved the way for the design and construction of readily programmable sequence-specific effectors for the precise manipulation of genomes.

[0006] Some aspects of the disclosure are based on the recognition that certain configurations of a nucleic acid programmable DNA binding protein (napDNAbp), for example CasX, CasY, Cpfl, C2cl, C2c2, C2c3, or Argonaute protein, and a cytidine deaminase domain fused by a linker are useful for efficiently deaminating target cytidine residues. Other aspects of this disclosure relate to the recognition that a nucleobase editing fusion protein with a cytidine deaminase domain fused to the N-terminus of a napDNAbp via a linker was capable of efficiently deaminating target nucleic acids in a double stranded DNA target molecule. See, for example, Examples 3 and 4 below, which demonstrate that the fusion proteins, which are also referred to herein as base editors, generate less indels and more efficiently deaminate target nucleic acids than other base editors, such as base editors without a UGI domain. Other aspects of this disclosure relate to the recognition that a nucleobase editing fusion protein with a cytidine deaminase domain fused to the N-terminus of napDNAbp via a linker perform base editing with higher efficiency and greatly improved product purity when the fusion protein is comprised of more than one UGI domain. See, for

example, Example 17, which demonstrates that a fusion protein (e.g., base editor) comprising two UGI domains generates less indels and more efficiently deaminates target nucleic acids than other base editors, such as those comprising one UGI domain.

[0007] In some embodiments, the fusion protein comprises: (i) a nucleic acid programmable DNA binding protein (napDNAbp); (ii) a cytidine deaminase domain; and (iii) a uracil glycosylase inhibitor (UGI) domain, where the napDNAbp is a CasX, CasY, Cpf1, C2c1, C2c2, C2c3, or Argonaute protein. In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a CasX protein. In some embodiments, the CasX protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 29 or 30. In some embodiments, the CasX protein comprises the amino acid sequence of SEQ ID NO: 29 or 30.

[0008] In some embodiments, the fusion protein comprises: (i) a nucleic acid programmable DNA binding protein (napDNAbp); (ii) a cytidine deaminase domain; (iii) a first uracil glycosylase inhibitor (UGI) domain; and (iv) a second uracil glycosylase inhibitor (UGI) domain, wherein the napDNAbp is a Cas9, dCas9, or Cas9 nickase protein. In some embodiments, the napDNAbp is a dCas9 protein. In some embodiments, the napDNAbp is a CasX, CasY, Cpf1, C2c1, C2c2, C2c3, or Argonaute protein. In some embodiments, the dCas9 protein is a *S. pyogenes* dCas9 (SpCas9d). In some embodiments, the dCas9 protein is a *S. pyogenes* dCas9 harboring a D10A mutation. In some embodiments, the dCas9 protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 6 or 7. In some embodiments, the dCas9 protein comprises the amino acid sequence of SEQ ID NO: 6 or 7. In some embodiments, the dCas9 protein is a *S. aureus* dCas9 (SaCas9d). In some embodiments, the dCas9 protein is a *S. aureus* dCas9 harboring a D10A mutation. In some embodiments, the dCas9 protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 33-36. In some embodiments, the dCas9 protein comprises the amino acid sequence of SEQ ID NO: 33-36.

[0009] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a CasY protein. In some embodiments, the CasY protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 31. In some embodiments, the CasY protein comprises the amino acid sequence of SEQ ID NO: 31.

[0010] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a Cpf1 or Cpf1 mutant protein. In some embodiments, the Cpf1 or Cpf1 mutant protein comprises an amino acid sequence that is at least 90% identical to any one of

SEQ ID NOs: 9-24. In some embodiments, the Cpfl or Cpf1 mutant protein comprises the amino acid sequence of any one of SEQ ID NOs: 9-24.

[0011] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a C2cl protein. In some embodiments, the C2cl protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 26. In some embodiments, the C2cl protein comprises the amino acid sequence of SEQ ID NO: 26.

[0012] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a C2c2 protein. In some embodiments, the C2c2 protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 27. In some embodiments, the C2c2 protein comprises the amino acid sequence of SEQ ID NO: 27.

[0013] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a C2c3 protein. In some embodiments, the C2c3 protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 28. In some embodiments, the C2c3 protein comprises the amino acid sequence of SEQ ID NO: 28.

[0014] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is an Argonaute protein. In some embodiments, the Argonaute protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 25. In some embodiments, the Argonaute protein comprises the amino acid sequence of SEQ ID NO: 25.

[0015] Some aspects of the disclosure are based on the recognition that fusion proteins provided herein are capable of generating one or more mutations *(e.g.,* a C to T mutation) without generating a large proportion of indels. In some embodiments, any of the fusion proteins *(e.g.,* base editing proteins) provided herein generate less than 10% indels. In some embodiments, any of the fusion proteins *(e.g.,* base editing proteins) provided herein generate less than 10%, 9%, 8%, 7%, 6%, 5.5%, 5%, 4.5%, 4%, 3.5%, 3%, 2.5%, 2%, 1.5%, 1%, 0.5%, or 0.1% indels.

[0016] In some embodiments, the fusion protein comprises a napDNAbp and an apolipoprotein B mRNA-editing complex 1 (APOBECl) deaminase domain, where the deaminase domain is fused to the N-terminus of the napDNAbp domain via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 604). In some embodiments, the napDNAbp comprises the amino acid sequence of any of the napDNAbp provided herein. In some embodiments, the deaminase is rat APOBECl (SEQ ID NO: 76). In some embodiments, the deaminase is human APOBECl (SEQ ID NO: 74). In some embodiments, the deaminase is pmCDAl (SEQ ID NO: 81). In some embodiments, the deaminase is human APOBEC3G (SEQ ID NO: 60). In some embodiments, the deaminase is a

human APOBEC3G variant of any one of (SEQ ID NOs: 82-84). In some embodiments, the fusion protein comprises a napDNAbp and an apolipoprotein B mRNA-editing complex 1 catalytic polypeptide-like 3G (APOBEC3G) deaminase domain, wherein the deaminase domain is fused to the N-terminus of the napDNAbp domain via a linker of any length or composition (e.g., an amino acid sequence, a peptide, a polymer, or a bond). In some embodiments, the linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 604). In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 605).

[0017] In some embodiments, the fusion protein comprises a napDNAbp and a cytidine deaminase 1 (CDA1) deaminase domain, wherein the deaminase domain is fused to the N-terminus of the napDNAbp domain via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 604). In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 605). In some embodiments, the napDNAbp comprises the amino acid sequence of any of the napDNAbps provided herein.

[0018] In some embodiments, the fusion protein comprises a napDNAbp and an activation-induced cytidine deaminase (AID) deaminase domain, where the deaminase domain is fused to the N-terminus of the napDNAbp domain via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 604). In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 605). In some embodiments, the napDNAbp comprises the amino acid sequence of any of the napDNAbps provided herein.

[0019] Some aspects of the disclosure are based on the recognition that certain configurations of a napDNAbp, and a cytidine deaminase domain fused by a linker are useful for efficiently deaminating target cytidine residues. Other aspects of this disclosure relate to the recognition that a nucleobase editing fusion protein with an apolipoprotein B mRNA-editing complex 1 (APOBEC1) deaminase domain fused to the N-terminus of a napDNAbp via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 604) was capable of efficiently deaminating target nucleic acids in a double stranded DNA target molecule. In some embodiments, the fusion protein comprises a napDNAbp domain and an apolipoprotein B mRNA-editing complex 1 (APOBEC1) deaminase domain, where the deaminase domain is fused to the N-terminus of the napDNAbp via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 604).

[0020] Some aspects of this disclosure provide strategies, systems, reagents, methods, and kits that are useful for the targeted editing of nucleic acids, including editing a single site within a subject's genome, *e.g.,* a human's genome. In some embodiments, fusion proteins of napDNAbp *(e.g.,* CasX, CasY, Cpfl, C2cl, C2c2, C2c3, or Argonaute protein) and deaminases or deaminase domains, are provided. In some embodiments, methods for targeted nucleic acid editing are provided. In some embodiments, reagents and kits for the generation of targeted nucleic acid editing proteins, *e.g.,* fusion proteins of napDNAbp and deaminases or deaminase domains, are provided.

[0021] Some aspects of this disclosure provide fusion proteins comprising a napDNAbp as provided herein that is fused to a second protein *(e.g.,* an enzymatic domain such as a cytidine deaminase domain), thus forming a fusion protein. In some embodiments, the second protein comprises an enzymatic domain, or a binding domain. In some embodiments, the enzymatic domain is a nuclease, a nickase, a recombinase, a deaminase, a methyltransferase, a methylase, an acetylase, an acetyltransferase, a transcriptional activator, or a transcriptional repressor domain. In some embodiments, the enzymatic domain is a nucleic acid editing domain. In some embodiments, the nucleic acid editing domain is a deaminase domain. In some embodiments, the deaminase is a cytosine deaminase or a cytidine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBEC 1 deaminase. In some embodiments, the deaminase is an APOBEC2 deaminase. In some embodiments, the deaminase is an APOBEC3 deaminase. In some embodiments, the deaminase is an APOBEC3A deaminase. In some embodiments, the deaminase is an APOBEC3B deaminase. In some embodiments, the deaminase is an APOBEC3C deaminase. In some embodiments, the deaminase is an APOBEC3D deaminase. In some embodiments, the deaminase is an APOBEC3E deaminase. In some embodiments, the deaminase is an APOBEC3F deaminase. In some embodiments, the deaminase is an APOBEC3G deaminase. In some embodiments, the deaminase is an APOBEC3H deaminase. In some embodiments, the deaminase is an APOBEC4 deaminase. In some embodiments, the deaminase is an activation-induced deaminase (AID). It should be appreciated that the deaminase may be from any suitable organism *(e.g.,* a human or a rat). In some embodiments, the deaminase is from a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase is rat APOBEC 1 (SEQ ID NO: 76). In some embodiments, the deaminase is human APOBEC 1 (SEQ ID NO: 74). In some embodiments, the deaminase is pmCDAl.

[0022] Some aspects of this disclosure provide fusion proteins comprising: (i) a CasX, CasY, Cpfl, C2cl, C2c2, C2c3, or Argonaute protein domain comprising the amino acid sequence of SEQ ID NO: 32; and (ii) an apolipoprotein B mRNA-editing complex 1 (APOBEC1) deaminase domain, wherein the deaminase domain is fused to the N-terminus of the napDNAbp via a linker comprising the amino acid sequence of SGSETPGTSESATPES (SEQ ID NO: 604). In some embodiments, the deaminase is rat APOBEC1 (SEQ ID NO: 76). In some embodiments, the deaminase is human APOBEC1 (SEQ ID NO: 74). In some embodiments, the fusion protein comprises the amino acid sequence of SEQ ID NO: 591. In some embodiments, the fusion protein comprises the amino acid sequence of SEQ ID NO: 5737. In some embodiments, the deaminase is pmCDAl (SEQ ID NO: 81). In some embodiments, the deaminase is human APOBEC3G (SEQ ID NO: 60). In some embodiments, the deaminase is a human APOBEC3G variant of any one of SEQ ID NOs: 82-84.

[0023] Other aspects of this disclosure relate to the recognition that fusion proteins comprising a deaminase domain, a napDNAbp domain and a uracil glycosylase inhibitor (UGI) domain demonstrate improved efficiency for deaminating target nucleotides in a nucleic acid molecule. Without wishing to be bound by any particular theory, cellular DNA-repair response to the presence of U:G heteroduplex DNA may be responsible for a decrease in nucleobase editing efficiency in cells. Uracil DNA glycosylase (UDG) catalyzes removal of U from DNA in cells, which may initiate base excision repair, with reversion of the U:G pair to a C:G pair as the most common outcome. As demonstrated herein, Uracil DNA Glycosylase Inhibitor (UGI) may inhibit human UDG activity. Without wishing to be bound by any particular theory, base excision repair may be inhibited by molecules that bind the single strand, block the edited base, inhibit UGI, inhibit base excision repair, protect the edited base, and/or promote "fixing" of the non-edited strand, *etc.* Thus, this disclosure contemplates fusion proteins comprising a napDNAbp-cytidine deaminase domain that is fused to a UGI domain.

[0024] Further aspects of this disclosure relate to the recognition that fusion proteins comprising a deaminase domain, a napDNAbp domain, and more than one uracil glycosylase inhibitor (UGI) domain *(e.g.,* one, two, three, four, five, or more UGI domains) demonstrate improved efficiency for deaminating target nucleotides in a nucleic acid molecule and/or improved nucleic acid product purity. Without wishing to be bound by any particular theory, the addition of a second UGI domain may substantially decrease the access of UDG to the G:U base editing intermediate, thereby improving the efficiency of the base editing.

[0025] Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of modifying a specific nucleotide base without generating a significant proportion of indels. An "indel", as used herein, refers to the insertion or deletion of a nucleotide base within a nucleic acid. Such insertions or deletions can lead to frame shift mutations within a coding region of a gene. In some embodiments, it is desirable to generate base editors that efficiently modify *(e.g.* mutate or deaminate) a specific nucleotide within a nucleic acid, without generating insertions or deletions *(i.e.,* indels) in the nucleic acid. In certain embodiments, any of the base editors provided herein are capable of generating a greater proportion of intended modifications *(e.g.,* point mutations or deaminations) versus indels.

[0026] In certain embodiments, any of the base editors provided herein are capable of generating a certain percentage of desired mutations. In some embodiments, the desired mutation is a C to T mutation. In some embodiments, the desired mutation is a C to A mutation, In some embodiments, the desired mutation is a C to G mutation. In some embodiments, any of the base editors provided herein are capable of generating at least 1% of desired mutations. In some embodiments, any of the base editors provided herein are capable of generating at least 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% of desired mutations.

[0027] Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of efficiently generating an intended mutation, such as a point mutation, in a nucleic acid *(e.g.* a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations, such as unintended point mutations.

[0028] In some embodiments, the deaminase domain of the fusion protein is fused to the N-terminus of the napDNAbp domain. In some embodiments, the UGI domain is fused to the C-terminus of the napDNAbp domain. In some embodiments, the napDNAbp and the nucleic acid editing domain are fused via a linker. In some embodiments, the napDNAbp domain and the UGI domain are fused via a linker. In some embodiments, a second UGI domain is fused to the C-terminus of a first UGI domain. In some embodiments, the first UGI domain and the second UGI domain are fused via a linker.

[0029] In certain embodiments, linkers may be used to link any of the peptides or peptide domains of the invention. The linker may be as simple as a covalent bond, or it may be a polymeric linker many atoms in length. In certain embodiments, the linker is a polpeptide or based on amino acids. In other embodiments, the linker is not peptide-like. In certain embodiments, the linker is a covalent bond *(e.g.,* a carbon-carbon bond, disulfide bond,

carbon-heteroatom bond, *etc.).* In certain embodiments, the linker is a carbon-nitrogen bond of an amide linkage. In certain embodiments, the linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or hetero aliphatic linker. In certain embodiments, the linker is polymeric *(e.g.,* polyethylene, polyethylene glycol, polyamide, polyester, *etc.).* In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In certain embodiments, the linker comprises an aminoalkanoic acid *(e.g.,* glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, *etc.).* In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain embodiments, the linker is based on a carbocyclic moiety *(e.g.,* cyclopentane, cyclohexane). In other embodiments, the linker comprises a polyethylene glycol moiety (PEG). In other embodiments, the linker comprises amino acids. In certain embodiments, the linker comprises a peptide. In certain embodiments, the linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. The linker may included funtionalized moieties to facilitate attachment of a nucleophile *(e.g.,* thiol, amino) from the peptide to the linker. Any electrophile may be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates.

[0030] In some embodiments, the linker comprises the amino acid sequence $(GGGGS)_n$ (SEQ ID NO: 607), $(G)_n$ (SEQ ID NO: 608), $(EAAAK)_n$ (SEQ ID NO: 609), $(GGS)_n$ (SEQ ID NO:610), $(SGGS)_n$ (SEQ ID NO: 606), SGSETPGTSESATPES (SEQ ID NO: 604), $(XP)_n$ (SEQ ID NO: 611), $SGGS(GGS)_n$ (SEQ ID NO: 612), SGGSSGGSSGS ETPGTS ESATPES SGGSSGGS (SEQ ID NO: 605), or any combination thereof, wherein n is independently an integer between 1 and 30, and X is any amino acid. In some embodiments, the linker comprises the amino acid sequence $(GGS)_n$ (SEQ ID NO: 610), wherein n is 1, 3, or 7. In some embodiments, the linker comprises the amino acid sequence $SGGS(GGS)_n$ (SEQ ID NO: 612), wherein n is 2. In some embodiments, the linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 604). In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGS ETPGTS ESATPES SGGSSGGS (SEQ ID NO: 605).

[0031] In some embodiments, the fusion protein comprises the structure [nucleic acid editing domain]-[optional linker sequence]-[napDNAbp]- [optional linker sequence]-[UGI]. In some embodiments, the fusion protein comprises the structure [nucleic acid editing domain]-[optional linker sequence]-[UGI]-[optional linker sequence]-[napDNAbp]; [UGI]-[optional

linker sequence]-[nucleic acid editing domain]-[optional linker sequence]-[napDNAbp]; [UGI]-[optional linker sequence]-[napDNAbp]-[optional linker sequence]-[nucleic acid editing domain]; [napDNAbp]-[optional linker sequence]-[UGI]-[optional linker sequence]-[nucleic acid editing domain]; [napDNAbp]-[optional linker sequence]-[nucleic acid editing domain]-[optional linker sequence]-[UGI]; or [nucleic acid editing domain]-[optional linker sequence]-[napDNAbp]-[optional linker sequence]-[first UGI]-[optional linker sequence]-[second UGI].

[0032] In some embodiments, the nucleic acid editing domain comprises a deaminase. In some embodiments, the nucleic acid editing domain comprises a deaminase. In some embodiments, the deaminase is a cytidine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBEC 1 deaminase, an APOBEC2 deaminase, an APOBEC3A deaminase, an APOBEC3B deaminase, an APOBEC3C deaminase, an APOBEC3D deaminase, an APOBEC3F deaminase, an APOBEC3G deaminase, an APOBEC3H deaminase, or an APOBEC4 deaminase. In some embodiments, the deaminase is an activation-induced deaminase (AID). In some embodiments, the deaminase is a cytidine deaminase 1 (CDA1). In some embodiments, the deaminase is a Lamprey CDA1 (pmCDAl) deaminase.

[0033] In some embodiments, the deaminase is from a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase is from a human. In some embodiments the deaminase is from a rat. In some embodiments, the deaminase is a rat APOBEC 1 deaminase comprising the amino acid sequence set forth in (SEQ ID NO: 76). In some embodiments, the deaminase is a human APOBEC 1 deaminase comprising the amino acid sequence set forth in (SEQ ID NO: 74). In some embodiments, the deaminase is pmCDAl (SEQ ID NO: 81). In some embodiments, the deaminase is human APOBEC 3G (SEQ ID NO: 60). In some embodiments, the deaminase is a human APOBEC3G variant of any one of (SEQ ID NOs: 82-84). In some embodiments, the deaminase is at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the amino acid sequences set forth in SEQ ID NOs: 49-84.

[0034] In some embodiments, the UGI domain comprises an amino acid sequence that is at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to SEQ ID NO: 134. In some

embodiments, the UGI domain comprises the amino acid sequence as set forth in SEQ ID NO: 134.

[0035] Some aspects of this disclosure provide complexes comprising a napDNAbp fusion protein as provided herein, and a guide RNA bound to the napDNAbp.

[0036] Some aspects of this disclosure provide methods of using the napDNAbp, fusion proteins, or complexes provided herein. For example, some aspects of this disclosure provide methods comprising contacting a DNA molecule (a) with a napDNAbp or a fusion protein as provided herein and with a guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence; or (b) with a napDNAbp, a napDNAbp fusion protein, or a napDNAbp or napDNAbp complex with a gRNA as provided herein.

[0037] Some aspects of this disclosure provide kits comprising a nucleic acid construct, comprising (a) a nucleotide sequence encoding a napDNAbp or a napDNAbp fusion protein as provided herein; and (b) a heterologous promoter that drives expression of the sequence of (a). In some embodiments, the kit further comprises an expression construct encoding a guide RNA backbone, wherein the construct comprises a cloning site positioned to allow the cloning of a nucleic acid sequence identical or complementary to a target sequence into the guide RNA backbone.

[0038] Some aspects of this disclosure provide polynucleotides encoding a napDNAbp of a fusion protein as provided herein. Some aspects of this disclosure provide vectors comprising such polynucleotides. In some embodiments, the vector comprises a heterologous promoter driving expression of polynucleotide.

[0039] Some aspects of this disclosure provide cells comprising a napDNAbp protein, a fusion protein, a nucleic acid molecule, and/or a vector as provided herein.

[0040] It should be appreciated that any of the fusion proteins provided herein that include a Cas9 domain *(e.g.* Cas9, nCas9, or dCas9) may be replaced with any of the napDNAbp provided herein, for example CasX, CasY, Cpfl, C2cl, C2c2, C2c3, or Argonaute protein.

[0041] The description of exemplary embodiments of the reporter systems above is provided for illustration purposes only and not meant to be limiting. Additional reporter systems, *e.g.,* variations of the exemplary systems described in detail above, are also embraced by this disclosure.

[0042] The summary above is meant to illustrate, in a non-limiting manner, some of the embodiments, advantages, features, and uses of the technology disclosed herein. Other

embodiments, advantages, features, and uses of the technology disclosed herein will be apparent from the Detailed Description, the Drawings, the Examples, and the Claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0043] **Figure 1** shows the deaminase activity of deaminases on single stranded DNA substrates. Single stranded DNA substrates using randomized PAM sequences (NNN PAM) were used as negative controls. Canonical PAM sequences used include the (NGG PAM).

[0044] **Figure 2** shows the activity of Cas9:deaminase fusion proteins on single stranded DNA substrates.

[0045] **Figure 3** illustrates double stranded DNA substrate binding by Cas9:deaminase:sgRNA complexes.

[0046] **Figure 4** illustrates a double stranded DNA deamination assay.

[0047] **Figure 5** demonstrates that Cas9 fusions can target positions 3-11 of double-stranded DNA target sequences (numbered according to the schematic in **Figure 5).** Upper Gel: 1 μM rAPOBECl-GGS-dCas9, 125 nM dsDNA, 1 equivalent sgRNA. Mid Gel: 1 μM rAPOBECl-$(GGS)_3$(SEQ ID NO: 610)-dCas9, 125 nM dsDNA, 1 equivalent sgRNA. Lower Gel: 1.85 μM rAPOBECl-XTEN-dCas9, 125 nM dsDNA, 1 equivalent sgRNA.

[0048] **Figure 6** demonstrates that the correct guide RNA, *e.g.*, the correct sgRNA, is required for deaminase activity.

[0049] **Figure 7** illustrates the mechanism of target DNA binding of *in vivo* target sequences by deaminase-dCas9:sgRNA complexes.

[0050] **Figure 8** shows successful deamination of exemplary disease-associated target sequences.

[0051] **Figure 9** shows *in vitro* C→T editing efficiencies using His6-rAPOBECl-XTEN-dCas9.

[0052] **Figure 10** shows C→T editing efficiencies in HEK293T cells is greatly enhanced by fusion with UGI.

[0053] **Figures 11A** to **11C** show NBE1 mediates specific, guide RNA-programmed C to U conversion *in vitro.* Figure 11A: Nucleobase editing strategy. DNA with a target C at a locus specified by a guide RNA is bound by dCas9, which mediates the local denaturation of the DNA substrate. Cytidine deamination by a tethered APOBEC1 enzyme converts the target C to U. The resulting G:U heteroduplex can be permanently converted to an A:T base pair following DNA replication or repair. If the U is in the template DNA strand, it will also result in an RNA transcript containing a G to A mutation following transcription. Figure 1IB:

Deamination assay showing an activity window of approximately five nucleotides. Following incubation of NBEl-sgRNA complexes with dsDNA substrates at 37 °C for 2 h, the 5' fluorophore-labeled DNA was isolated and incubated with USER enzyme (uracil DNA glycosylase and endonuclease VIII) at 37 °C for 1 h to induce DNA cleavage at the site of any uracils. The resulting DNA was resolved on a denaturing polyacrylamide gel, and any fluorophore-linked strands were visualized. Each lane is labeled according to the position of the target C within the protospacer, or with "-" if no target C is present, counting the base distal from the PAM as position 1. Figure 11C: Deaminase assay showing the sequence specificity and sgRNA-dependence of NBE1. The DNA substrate with a target C at position 7 was incubated with NBE1 as in Figure 1IB with either the correct sgRNA, a mismatched sgRNA, or no sgRNA. No C to U editing is observed with the mismatched sgRNA or with no sgRNA. The positive control sample contains a DNA sequence with a U synthetically incorporated at position 7.

[0054] **Figures 12A** to **12B** show effects of sequence context and target C position on nucleobase editing efficiency *in vitro.* Figure 12A: Effect of changing the sequence surrounding the target C on editing efficiency *in vitro.* The deamination yield of 80% of targeted strands (40% of total sequencing reads from both strands) for C_7 in the protospacer sequence 5'-TTATTTCGTGGATTTATTTA-3'(SEQ ID NO: 591) was defined as 1.0, and the relative deamination efficiencies of substrates containing all possible single-base mutations at positions 1-6 and 8-13 are shown. Values and error bars reflect the mean and standard deviation of two or more independent biological replicates performed on different days. Figure 12B: Positional effect of each NC motif on editing efficiency *in vitro.* Each NC target motif was varied from positions 1 to 8 within the protospacer as indicated in the sequences shown on the right (the PAM shown in red, the protospacer plus one base 5' to the protospacer are also shown). The percentage of total sequence reads containing T at each of the numbered target C positions following incubation with NBE1 is shown in the graph. Note that the maximum possible deamination yield *in vitro* is 50% of total sequencing reads (100% of targeted strands). Values and error bars reflect the mean and standard deviation of two or three independent biological replicates performed on different days. Figure 12B depicts SEQ ID NOs: 619 through 626from top to bottom, respectively.

[0055] **Figures 13A** to **13C** show nucleobase editing in human cells. Figure 13A: Protospacer and PAM sequences of the six mammalian cell genomic loci targeted by nucleobase editors. Target Cs are indicated with subscripted numbers corresponding to their positions within the protospacer. Figure 13A depicts SEQ ID NOs: 127 through 132 from top

to bottom, respectively. Figure 13B: HEK293T cells were transfected with plasmids expressing NBE1, NBE2, or NBE3 and an appropriate sgRNA. Three days after transfection, genomic DNA was extracted and analyzed by high-throughput DNA sequencing at the six loci. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with Ts at the target positions indicated, are shown for NBE1, NBE2, and NBE3 at all six genomic loci, and for wt Cas9 with a donor HDR template at three of the six sites (EMX1, HEK293 site 3, and HEK293 site 4). Values and error bars reflect the mean and standard deviation of three independent biological replicates performed on different days. Figure 13C: Frequency of indel formation, calculated as described in the Methods, is shown following treatment of HEK293T cells with NBE2 and NBE3 for all six genomic loci, or with wt Cas9 and a single-stranded DNA template for HDR at three of the six sites (EMX1, HEK293 site 3, and HEK293 site 4). Values reflect the mean of at least three independent biological replicates performed on different days.

[0056] **Figures 14A** to **14C** show NBE2- and NBE3-mediated correction of three disease-relevant mutations in mammalian cells. For each site, the sequence of the protospacer is indicated to the right of the name of the mutation, with the PAM and the base responsible for the mutation indicated in bold with a subscripted number corresponding to its position within the protospacer. The amino acid sequence above each disease-associated allele is shown, together with the corrected amino acid sequence following nucleobase editing in red. Underneath each sequence are the percentages of total sequencing reads with the corresponding base. Cells were nucleofected with plasmids encoding NBE2 or NBE3 and an appropriate sgRNA. Two days after nucleofection, genomic DNA was extracted and analyzed by HTS to assess pathogenic mutation correction. Figure 14A: The Alzheimer's disease-associated *APOE4* allele is converted to *APOE3'* in mouse astrocytes by NBE3 in 11% of total reads (44% of nucleofected astrocytes). Two nearby Cs are also converted to Ts, but with no change to the predicted sequence of the resulting protein (SEQ ID NO: 627). Figure 14B The cancer-associated p53 N239D mutation is corrected by NBE2 in 11% of treated human lymphoma cells (12% of nucleofected cells) that are heterozygous for the mutation (SEQ ID NO: 628). Figure 14C The p53 Y163C mutation is corrected by NBE3 in 7.6% of nucleofected human breast cancer cells (SEQ ID NO: 629).

[0057] **Figures 15A** to **15D** show effects of deaminase-dCas9 linker length and composition on nucleobase editing. Gel-based deaminase assay showing the deamination window of nucleobase editors with deaminase-Cas9 linkers of GGS (Figure 15A), $(GGS)_3$ (SEQ ID NO: 610) (Figure 15B), XTEN (Figure 15C), or $(GGS)_7$ (SEQ ID NO: 610) (Figure 15D).

Following incubation of 1.85 μM editor-sgRNA complexes with 125 iiM dsDNA substrates at 37 °C for 2 h, the dye-conjugated DNA was isolated and incubated with USER enzyme (uracil DNA glycosylase and endonuclease VIII) at 37 °C for an additional hour to cleave the DNA backbone at the site of any uracils. The resulting DNA was resolved on a denaturing polyacrylamide gel, and the dye-conjugated strand was imaged. Each lane is numbered according to the position of the target C within the protospacer, or with - if no target C is present. 8U is a positive control sequence with a U synthetically incorporated at position 8.

[0058] **Figures 16A** to **16B** show NBE1 is capable of correcting disease-relevant mutations *in vitro*. Figure 16A: Protospacer and PAM sequences of seven disease-relevant mutations. The disease-associated target C in each case is indicated with a subscripted number reflecting its position within the protospacer. For all mutations except both *APOE4* SNPs, the target C resides in the template (non-coding) strand. Figure 16A depicts SEQ ID NOs: 631 through 636 from top to bottom, respectively. Figure 16B: Deaminase assay showing each dsDNA oligonucleotide before (-) and after (+) incubation with NBE1, DNA isolation, and incubation with USER enzymes to cleave DNA at positions containing U. Positive control lanes from incubation of synthetic oligonucleotides containing U at various positions within the protospacer with USER enzymes are shown with the corresponding number indicating the position of the U.

[0059] **Figure 17** shows processivity of NBE1. The protospacer and PAM of a 60-mer DNA oligonucleotide containing eight consecutive Cs is shown at the top. The oligonucleotide (125 iiM) was incubated with NBE1 (2 μM) for 2 h at 37 °C. The DNA was isolated and analyzed by high-throughput sequencing. Shown are the percent of total reads for the most frequent nine sequences observed. The vast majority of edited strands (>93%) have more than one C converted to T. This figure depicts SEQ ID NO: 309.

[0060] **Figures 18A** to **18H** show the effect of fusing UGI to NBEl to generate NBE2. Figure 18A: Protospacer and PAM sequences of the six mammalian cell genomic loci targeted with nucleobase editors. Editable Cs are indicated with labels corresponding to their positions within the protospacer. Figure 18A depicts SEQ ID NOs: 127 through 132 from top to bottom, respectively. Figures 18B to 18G: HEK293T cells were transfected with plasmids expressing NBEl, NBE2, or NBEl and UGI, and an appropriate sgRNA. Three days after transfection, genomic DNA was extracted and analyzed by high-throughput DNA sequencing at the six loci. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with Ts at the target positions indicated, are shown for NBEl, NBEl and UGI, and NBE2 at all six genomic loci. Figure 18H: C to T mutation rates at 510 Cs

surrounding the protospacers of interest for NBEl, NBEl plus UGI on a separate plasmid, NBE2. and untreated cells are shown. The data show the results of 3,000,000 DNA sequencing reads from 1.5x106 cells. Values reflect the mean of at least two biological experiments conducted on different days.

[0061] **Figure 19** shows nucleobase editing efficiencies of NBE2 in U20S and HEK293T cells. Cellular C to T conversion percentages by NBE2 are shown for each of the six targeted genomic loci in HEK293T cells and U20S cells. HEK293T cells were transfected using lipofectamine 2000, and U20S cells were nucleofected. U20S nucleofection efficiency was 74%. Three days after plasmid delivery, genomic DNA was extracted and analyzed for nucleobase editing at the six genomic loci by HTS. Values and error bars reflect the mean and standard deviation of at least two biological experiments done on different days.

[0062] **Figure 20** shows nucleobase editing persists over multiple cell divisions. Cellular C to T conversion percentages by NBE2 are displayed at two genomic loci in HEK293T cells before and after passaging the cells. HEK293T cells were transfected using Lipofectamine 2000. Three days post transfection, the cells were harvested and split in half. One half was subjected to HTS analysis, and the other half was allowed to propagate for approximately five cell divisions, then harvested and subjected to HTS analysis.

[0063] **Figure 21** shows genetic variants from ClinVar that can be corrected in principle by nucleobase editing. The NCBI ClinVar database of human genetic variations and their corresponding phenotypes[68] was searched for genetic diseases that can be corrected by current nucleobase editing technologies. The results were filtered by imposing the successive restrictions listed on the left. The x-axis shows the number of occurrences satisfying that restriction and all above restrictions on a logarithmic scale.

[0064] **Figure 22** shows *in vitro* identification of editable Cs in six genomic loci. Synthetic 80-mers with sequences matching six different genomic sites were incubated with NBEl then analyzed for nucleobase editing *via* HTS. For each site, the sequence of the protospacer is indicated to the right of the name of the site, with the PAM highlighted in red. Underneath each sequence are the percentages of total DNA sequencing reads with the corresponding base. A target C was considered as "editable" if the *in vitro* conversion efficiency is >10%. Note that maximum yields are 50% of total DNA sequencing reads since the non-targeted strand is not a substrate for nucleobase editing. This figure depicts SEQ ID NOs: 127 through 132 from top to bottom, respectively.

[0065] **Figure 23** shows activities of NBEl, NBE2, and NBE3 at EMX1 off-targets. HEK293T cells were transfected with plasmids expressing NBEl, NBE2, or NBE3 and a

sgRNA matching the EMX1 sequence using Lipofectamine 2000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target loci, plus the top ten known Cas9 off-target loci for the EMX1 sgRNA, as previously determined using the GUIDE-seq method[55]. EMX1 off-target 5 locus did not amplify and is not shown. Sequences of the on-target and off-target protospacers and protospacer adjacent motifs (PAMs) are displayed. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with T at each position of an original C within the protospacer, are shown for NBE1, NBE2, and NBE3. On the far right are displayed the total number of sequencing reads reported for each sequence. This figure depicts SEQ ID NOs: 127, and 637 through 645 from top to bottom, respectively.

[0066] **Figure 24** shows activities of NBE1, NBE2, and NBE3 at FANCF off-targets. HEK293T cells were transfected with plasmids expressing NBE1, NBE2, or NBE3 and a sgRNA matching the FANCF sequence using Lipofectamine 2000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target loci, plus all of the known Cas9 off-target loci for the FANCF sgRNA, as previously determined using the GUIDE-seq method[55]. Sequences of the on-target and off-target protospacers and protospacer adjacent motifs (PAMs) are displayed. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with T at each position of an original C within the protospacer, are shown for NBE1, NBE2, and NBE3. On the far right are displayed the total number of sequencing reads reported for each sequence. This figure depicts SEQ ID NOs: 128 and 646 through 653 from top to bottom, respectively.

[0067] **Figure 25** shows activities of NBE1, NBE2, and NBE3 at HEK293 site 2 off-targets. HEK293T cells were transfected with plasmids expressing NBE1, NBE2, or NBE3 and a sgRNA matching the HEK293 site 2 sequence using Lipofectamine 2000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target loci, plus all of the known Cas9 off-target loci for the HEK293 site 2 sgRNA, as previously determined using the GUIDE-seq method[55]. Sequences of the on-target and off-target protospacers and protospacer adjacent motifs (PAMs) are displayed. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with T at each position of an original C within the protospacer, are shown for NBE1, NBE2, and NBE3. On the far right are displayed the total number of sequencing reads reported for each sequence. This figure depicts SEQ ID NOs: 129, 654, and 655 from top to bottom, respectively.

[0068] **Figure 26** shows activities of NBE1, NBE2, and NBE3 at HEK293 site 3 off-targets. HEK293T cells were transfected with plasmids expressing NBE1, NBE2, or NBE3 and a sgRNA matching the HEK293 site 3 sequence using Lipofectamine 2000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target loci, plus all of the known Cas9 off-target loci for the HEK293 site 3 sgRNA, as previously determined using the GUIDE-seq method.[55] Sequences of the on-target and off-target protospacers and protospacer adjacent motifs (PAMs) are displayed. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with T at each position of an original C within the protospacer, are shown for NBE1, NBE2, and NBE3. On the far right are displayed the total number of sequencing reads reported for each sequence. This figure depicts SEQ ID NOs: 130 and 656 through 660 from top to bottom, respectively.

[0069] **Figure 27** shows activities of NBE1, NBE2, and NBE3 at HEK293 site 4 off-targets. HEK293T cells were transfected with plasmids expressing NBE1, NBE2, or NBE3 and a sgRNA matching the HEK293 site 4 sequence using Lipofectamine 2000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target loci, plus the top ten known Cas9 off-target loci for the HEK293 site 4 sgRNA, as previously determined using the GUIDE-seq method.[55] Sequences of the on-target and off-target protospacers and protospacer adjacent motifs (PAMs) are displayed. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with T at each position of an original C within the protospacer, are shown for NBE1, NBE2, and NBE3. On the far right are displayed the total number of sequencing reads reported for each sequence. This figure depicts SEQ ID NOs: 131 and 661 through 670 from top to bottom, respectively.

[0070] **Figure 28** shows non-target C mutation rates. Shown here are the C to T mutation rates at 2,500 distinct cytosines surrounding the six on-target and 34 off-target loci tested, representing a total of 14,700,000 sequence reads derived from approximately 1.8x106 cells.

[0071] **Figures 29A** to **29C** show base editing in human cells. Figure 29A shows possible base editing outcomes in mammalian cells. Initial editing resulted in a U:G mismatch. Recognition and excision of the U by uracil DNA glycosylase (UDG) initiated base excision repair (BER), which lead to reversion to the C:G starting state. BER was impeded by BE2 and BE3, which inhibited UDG. The U:G mismatch was also processed by mismatch repair (MMR), which preferentially repaired the nicked strand of a mismatch. BE3 nicked the non-edited strand containing the G, favoring resolution of the U:G mismatch to the desired U:A or

T:A outcome. Figure 29B shows HEK293T cells treated as described in the Materials and Methods in the Examples below. The percentage of total DNA sequencing read with Ts at the target positions indicated show treatment with BE1, BE2, or BE3, or for treatment with wt Cas9 with a donor HDR template. Figure 29C shows frequency of indel formation following the treatment in Figure 29B. Values are listed in Figure 34. For Figures 29B and 29C, values and error bars reflect the mean and s.d. of three independent biological replicates performed on different days.

[0072] **Figures 30A** to **30B** show BE3-mediated correction of two disease-relevant mutations in mammalian cells. The sequence of the protospacer is shown to the right of the mutation, with the PAM and the target base in red with a subscripted number indicating its position within the protospacer. Underneath each sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Materials and Methods. Figure 30A shows the Alzheimer's disease-associated APOE4 allele converted to APOE3r in mouse astrocytes by BE3 in 74.9% of total reads. Two nearby Cs were also converted to Ts, but with no change to the predicted sequence of the resulting protein. Identical treatment of these cells with wt Cas9 and donor ssDNA results in only 0.3% correction, with 26.1% indel formation. This figure depicts SEQ ID NOs: 671 and 627. Figure 30B shows the cancer associated p53 Y163C mutation corrected by BE3 in 7.6% of nucleofected human breast cancer cells with 0.7% indel formation. Identical treatment of these cells with wt Cas9 and donor ssDNA results in no mutation correction with 6.1% indel formation. This figure depicts SEQ ID NOs: 672 and 629.

[0073] **Figure 31** shows activities of BE1, BE2, and BE3 at HEK293 site 2 off-targets. HEK293T cells were transfected with plasmids expressing BE1, BE2, or BE3 and a sgRNA matching the HEK293 site 2 sequence using Lipofectamine 2000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target loci, plus all of the known Cas9 and dCas9 off-target loci for the HEK293 site 2 sgRNA, as previously determined by Joung and coworkers using the GUIDE-seq method (63), and Adli and coworkers using chromatin immunoprecipitation high-throughput sequencing (ChlP-seq) experiments (18). Sequences of the on-target and off-target protospacers and protospacer adjacent motifs (PAMs) are displayed. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with T at each position of an original C within the protospacer, are shown for BE1, BE2, and BE3. On the far right are displayed the total number of sequencing reads

reported, and the ChlP-seq signal intensity reported for each sequence. This figure depicts SEQ ID NOs: 129, 654, 655 and 673 to 677 from top to bottom, respectively.

[0074] **Figure 32** shows activities of BEl, BE2, and BE3 at HEK293 site 3 off-targets. HEK293T cells were transfected with plasmids expressing BEl, BE2, or BE3 and a sgRNA matching the HEK293 site 3 sequence using Lipofectamine 2000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target loci, plus all of the known Cas9 off-target loci and the top five known dCas9 off-target loci for the HEK293 site 3 sgRNA, as previously determined by Joung and coworkers using the GUIDE-seq method[54], and using chromatin immunoprecipitation high-throughput sequencing (ChlP-seq) experiments[61]. Sequences of the on-target and off-target protospacers and protospacer adjacent motifs (PAMs) are displayed. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with T at each position of an original C within the protospacer, are shown for BEl, BE2, and BE3. On the far right are displayed the total number of sequencing reads reported, and the ChlP-seq signal intensity reported for each sequence. This figure depicts SEQ ID NOs: 130, 656 to 660 and 678-682 from top to bottom, respectively.

[0075] **Figure 33** shows activities of BEl, BE2, and BE3 at HEK293 site 4 off-targets. HEK293T cells were transfected with plasmids expressing BEl, BE2, or BE3 and a sgRNA matching the HEK293 site 4 sequence using Lipofectamine 2000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target loci, plus the top ten known Cas9 off-target loci and the top five known dCas9 off-target loci for the HEK293 site 4 sgRNA, as previously determined using the GUIDE-seq method[54], and using chromatin immunoprecipitation high-throughput sequencing (ChlP-seq) experiments[61]. Sequences of the on-target and off-target protospacers and protospacer adjacent motifs (PAMs) are displayed. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with T at each position of an original C within the protospacer, are shown for BEl, BE2, and BE3. On the far right are displayed the total number of sequencing reads reported, and the ChlP-seq signal intensity reported for each sequence. This figure depicts SEQ ID NOs: 131, 661 to 670, 683 and 684 from top to bottom, respectively.

[0076] **Figure 34** shows mutation rates of non-protospacer bases following BE3-mediated correction of the Alzheimer's disease-associated *APOE4* allele to *APOE3r* in mouse astrocytes. The DNA sequence of the 50 bases on either side of the protospacer from Figure 30A and Figure 34B is shown with each base's position relative to the protospacer. The side of

the protospacer distal to the PAM is designated with positive numbers, while the side that includes the PAM is designated with negative numbers, with the PAM. Underneath each sequence are the percentages of total DNA sequencing reads with the corresponding base for untreated cells, for cells treated with BE3 and an sgRNA targeting the *APOE4* C158R mutation, or for cells treated with BE3 and an sgRNA targeting the *VEGFA* locus. Neither BE3-treated sample resulted in mutation rates above those of untreated controls. This figure depicts SEQ ID NOs: 685 to 688 from top to bottom, respectively.

[0077] **Figure 35** shows mutation rates of non-protospacer bases following BE3-mediated correction of the cancer-associated p53 Y163C mutation in HCC1954 human cells. The DNA sequence of the 50 bases on either side of the protospacer from Figure 30B and Figure 39Bis shown with each base's position relative to the protospacer. The side of the protospacer distal to the PAM is designated with positive numbers, while the side that includes the PAM is designated with negative numbers, with the PAM. Underneath each sequence are the percentages of total sequencing reads with the corresponding base for untreated cells, for cells treated with BE3 and an sgRNA targeting the *TP53* Y163C mutation, or for cells treated with BE3 and an sgRNA targeting the *VEGFA* locus. Neither BE3-treated sample resulted in mutational rates above those of untreated controls. This figure depicts SEQ ID NOs: 689 to692 from top to bottom, respectively.

[0078] **Figures 36A** to **36F** show the effects of deaminase, linker length, and linker composition on base editing. Figure 36A shows a gel-based deaminase assay showing activity of rAPOBECl, pmCDAl, hAID, hAPOBEC3G, rAPOBECl-GGS-dCas9, rAPOBECl-$(GGS)_3$(SEQ ID NO: 610)-dCas9, and dCas9-$(GGS)_3$(SEQ ID NO: 610)-rAPOBECl on ssDNA. Enzymes were expressed in a mammalian cell lysate-derived *in vitro* transcription-translation system and incubated with 1.8 μM dye-conjugated ssDNA and USER enzyme (uracil DNA glycosylase and endonuclease VIII) at 37 °C for 2 hours. The resulting DNA was resolved on a denaturing polyacrylamide gel and imaged. The positive control is a sequence with a U synthetically incorporated at the same position as the target C. Figure 36B shows coomassie-stained denaturing PAGE gel of the expressed and purified proteins used in Figures 36C to 36F. Figures 36C to 36F show gel-based deaminase assay showing the deamination window of base editors with deaminase-Cas9 linkers of GGS (Figure 36C), $(GGS)_3$ (SEQ ID NO: 610) (Figure 36D), XTEN (Figure 36E), or $(GGS)_7$ (SEQ ID NO: 610) (Figure 36F). Following incubation of 1.85 μM deaminase-dCas9 fusions complexed with sgRNA with 125 nM dsDNA substrates at 37 °C for 2 hours, the dye-conjugated DNA was isolated and incubated with USER enzyme at 37 °C for 1 hour to cleave the DNA backbone at the site of

any uracils. The resulting DNA was resolved on a denaturing polyacrylamide gel, and the dye-conjugated strand was imaged. Each lane is numbered according to the position of the target C within the protospacer, or with - if no target C is present. 8U is a positive control sequence with a U synthetically incorporated at position 8. .

[0079] **Figures 37A** to **37C** show BEl base editing efficiencies are dramatically decreased in mammalian cells. Figure 37A Protospacer and PAM sequences of the six mammalian cell genomic loci targeted by base editors. Target Cs are indicated in red with subscripted numbers corresponding to their positions within the protospacer. Figure 37B shows synthetic 80-mers with sequences matching six different genomic sites were incubated with BEl then analyzed for base editing by HTS. For each site, the sequence of the protospacer is indicated to the right of the name of the site, with the PAM. Underneath each sequence are the percentages of total DNA sequencing reads with the corresponding base. We considered a target C as "editable" if the *in vitro* conversion efficiency is >10%. Note that maximum yields are 50% of total DNA sequencing reads since the non-targeted strand is unaffected by BEl. Values are shown from a single experiment. Figure 37C shows HEK293T cells were transfected with plasmids expressing BEl and an appropriate sgRNA. Three days after transfection, genomic DNA was extracted and analyzed by high-throughput DNA sequencing at the six loci. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with Ts at the target positions indicated, are shown for BEl at all six genomic loci. Values and error bars of all data from HEK293T cells reflect the mean and standard deviation of three independent biological replicates performed on different days. Figure 37A depicts SEQ ID NOs: 127 to 132 from top to bottom, respectively. Figure 37B depicts SEQ ID NOs: 127 to 132 from top to bottom, respectively.

[0080] **Figure 38** shows base editing persists over multiple cell divisions. Cellular C to T conversion percentages by BE2 and BE3 are shown for HEK293 sites 3 and 4 in HEK293T cells before and after passaging the cells. HEK293T cells were nucleofected with plasmids expressing BE2 or BE3 and an sgRNA targeting HEK293 site 3 or 4. Three days after nucleofection, the cells were harvested and split in half. One half was subjected to HTS analysis, and the other half was allowed to propagate for approximately five cell divisions, then harvested and subjected to HTS analysis. Values and error bars reflect the mean and standard deviation of at least two biological experiments.

[0081] **Figures 39A** to **39C** show non-target C/G mutation rates. Shown here are the C to T and G to A mutation rates at 2,500 distinct cytosines and guanines surrounding the six on-target and 34 off-target loci tested, representing a total of 14,700,000 sequence reads derived

from approximately $1.8xl0^6$ cells. Figures 39A and 39B show cellular non-target C to T and G to A conversion percentages by BE1, BE2, and BE3 are plotted individually against their positions relative to a protospacer for all 2,500 cytosines/guanines. The side of the protospacer distal to the PAM is designated with positive numbers, while the side that includes the PAM is designated with negative numbers. Figure 39C shows average non-target cellular C to T and G to A conversion percentages by BE1, BE2, and BE3 are shown, as well as the highest and lowest individual conversion percentages.

[0082] **Figures 40A** to **40B** show additional data sets of BE3-mediated correction of two disease-relevant mutations in mammalian cells. For each site, the sequence of the protospacer is indicated to the right of the name of the mutation, with the PAM and the base responsible for the mutation indicated in red bold with a subscripted number corresponding to its position within the protospacer. The amino acid sequence above each disease-associated allele is shown, together with the corrected amino acid sequence following base editing. Underneath each sequence are the percentages of total sequencing reads with the corresponding base. Cells were nucleofected with plasmids encoding BE3 and an appropriate sgRNA. Two days after nucleofection, genomic DNA was extracted from the nucleofected cells and analyzed by HTS to assess pathogenic mutation correction. Figure 40A shows the Alzheimer's disease-associated *APOE4* allele is converted to *APOE3r* in mouse astrocytes by BE3 in 58.3% of total reads only when treated with the correct sgRNA. Two nearby Cs are also converted to Ts, but with no change to the predicted sequence of the resulting protein. Identical treatment of these cells with wt Cas9 and donor ssDNA results in 0.2% correction, with 26.7% indel formation. Figure 40B shows the cancer-associated p53 Y163C mutation is corrected by BE3 in 3.3% of nucleofected human breast cancer cells only when treated with the correct sgRNA. Identical treatment of these cells with wt Cas9 and donor ssDNA results in no detectable mutation correction with 8.0% indel formation. Figures 40A to 40B depict SEQ ID NOs: 671, 627, 672 and 629.

[0083] **Figure 41** shows a schematic representation of an exemplary USER (Uracil-Specific Excision Reagent) Enzyme-based assay, which may be used to test the activity of various deaminases on single-stranded DNA (ssDNA) substrates.

[0084] **Figure 42** is a schematic of the pmCDA-nCas9-UGI-NLS construct and its activity at the HeK-3 site relative to the base editor (rAPOBECl) and the negative control (untreated). This figure depicts SEQ ID NO: 693.

[0085] **Figure 43** is a schematic of the pmCDAl-XTEN-nCas9-UGI-NLS construct and its activity at the HeK-3 site relative to the base editor (rAPOBECl) and the negative control (untreated). This figure depicts SEQ ID NO: 694.

[0086] **Figure 44** shows the percent of total sequencing reads with target C converted to T using cytidine deaminases (CDA) or APOBEC.

[0087] **Figure 45** shows the percent of total sequencing reads with target C converted to A using deaminases (CDA) or APOBEC.

[0088] **Figure 46** shows the percent of total sequencing reads with target C converted to G using deaminases (CDA) or APOBEC.

[0089] **Figure 47** is a schematic of the huAPOBEC3G-XTEN-nCas9-UGI-NLS construct and its activity at the HeK-2 site relative to a mutated form (huAPOBEC3G*(D316R_D317R)-XTEN-nCas9-UGI-NLS, the base editor (rAPOBECl) and the negative control (untreated). This figure depicts SEQ ID NO: 695.

[0090] **Figure 48** shows the schematic of the LacZ construct used in the selection assay of Example 7.

[0091] **Figure 49** shows reversion data from different plasmids and constructs.

[0092] **Figure 50** shows the verification of lacZ reversion and the purification of reverted clones.

[0093] **Figure 51** is a schematic depicting a deamination selection plasmid used in Example 7.

[0094] **Figure 52** shows the results of a chloramphenicol reversion assay (pmCDAl fusion).

[0095] **Figures 53A** to **53B** demonstrated DNA correction induction of two constructs.

[0096] **Figure 54** shows the results of a chloramphenicol reversion assay (huAPOBEC3G fusion).

[0097] **Figure 55** shows the activities of BE3 and HF-BE3 at EMX1 off-targets. The sequences, from top to bottom, correspond to SEQ ID NOs: 127 and 637-645.

[0098] **Figure 56** shows on-target base editing efficiencies of BE3 and HF-BE3.

[0099] **Figure 57** is a graph demonstrating that mutations affect cytidine deamination with varying degrees. Combinations of mutations that each slightly impairs catalysis allow selective deamination at one position over others. The FANCF site was GGAATC$_6$C$_7$C$_8$TTCiiTGCAGCACCTGG (SEQ ID NO: 128).

[00100] **Figure 58** is a schematic depicting next generation base editors.

[00101] **Figure 59** is a schematic illustrating new base editors made from Cas9 variants.

[00102] **Figure 60** shows the base-edited percentage of different NGA PAM sites.

[00103] **Figure 61** shows the base-edited percentage of cytidines using NGCG PAM EMX (VRER BE3) and the CiTC$_3$C$_4$C$_5$ATC$_8$ACi$_0$ATCAACCGGT (SEQ ID NO: 696) spacer.

[00104] **Figure 62** shows the based-edited percentages resulting from different NNGRRT PAM sites.

[00105] **Figure 63** shows the based-edited percentages resulting from different NNHRRT PAM sites.

[00106] **Figures 64A** to **64C** show the base-edited percentages resulting from different TTTN PAM sites using Cpfl BE2. The spacers used were:
TTTCCTC$_3$C$_4$C$_5$C$_6$C$_7$C$_8$C9AC$_{11}$AGGTAGAACAT (Figure 64A, SEQ ID NO: 697),
TTTCCiC$_2$TC$_4$TGTC$_8$C$_9$ACiiACCCTCATCCTG (Figure 64B, SEQ ID NO: 698), and
TTTCCiC$_2$C$_3$AGTC$_7$C$_8$TCioCiiACi$_3$ACi$_5$Ci$_6$Ci$_7$TGAAAC (Figure 64C, SEQ ID NO: 699).

[00107] **Figure 65** is a schematic depicting selective deamination as achieved through kinetic modulation of cytidine deaminase point mutagenesis.

[00108] **Figure 66** is a graph showing the effect of various mutations on the deamination window probed in cell culture with multiple cytidines in the spacer. The spacer used was: TGC3C$_4$C$_5$C6TC$_8$C$_9$CioTCi$_2$Ci3Ci$_4$TGGCCC (SEQ ID NO: 700).

[00109] **Figure 67** is a graph showing the effect of various mutations on the deamination window probed in cell culture with multiple cytidines in the spacer. The spacer used was: AGAGC$_5$C$_6$C$_7$C$_8$C9CioCiiTCi$_3$AAAGAGA (SEQ ID NO: 701).

[00110] **Figure 68** is a graph showing the effect of various mutations on the FANCF site with a limited number of cytidines. The spacer used was:
GGAATC$_6$C$_7$C$_8$TTCiiTGCAGCACCTGG (SEQ ID NO: 128). Note that the triple mutant (W90Y, R126E, R132E) preferentially edits the cytidine at the sixth position.

[00111] **Figure 69** is a graph showing the effect of various mutations on the HEK3 site with a limited number of cytidines. The spacer used was: GGCC$_4$C$_5$AGACTGAGCACGTGATGG (SEQ ID NO: 702). Note that the double and triple mutants preferentially edit the cytidine at the fifth position over the cytidine in the fourth position.

[00112] **Figure 70** is a graph showing the effect of various mutations on the EMX1 site with a limited number of cytidines. The spacer used was:
GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG (SEQ ID NO: 703). Note that the triple mutant only edits the cytidine at the fifth position, not the sixth.

[00113] **Figure 71** is a graph showing the effect of various mutations on the HEK2 site with a limited number of cytidines. The spacer used was:
GAAC$_4$AC$_6$AAAGCATAGACTGCGGG (SEQ ID NO: 704).

[00114] **Figure 72** shows on-target base editing efficiencies of BE3 and BE3 comprising mutations W90Y R132E in immortalized astrocytes.

[00115] **Figure 73** depicts a schematic of three Cpfl fusion constructs.

[00116] **Figures 74** shows a comparison of plasmid delivery of BE3 and HF-BE3 (EMX1, FANCF, and RNF2).

[00117] **Figure 75** shows a comparison of plasmid delivery of BE3 and HF-BE3 (HEK3 and HEK 4).

[00118] **Figure 76** shows off-target editing of EMX-1 at all 10 sites. This figure depicts SEQ ID NOs: 127 and 637-645

[00119] **Figure 77** shows deaminase protein lipofection to HEK cells using a GAGTCCGAGCAGAAGAAGAAG (SEQ ID NO: 705) spacer. The EMX-1 on-target and EMX-1 off target site 2 were examined.

[00120] **Figure 78** shows deaminase protein lipofection to HEK cells using a GGAATCCCTTCTGCAGCACCTGG (SEQ ID NO: 706) spacer. The FANCF on target and FANCF off target site 1 were examined.

[00121] **Figure 79** shows deaminase protein lipofection to HEK cells using a GGCCCAGACTGAGCACGTGA (SEQ ID NO: 707) spacer. The HEK-3 on target site was examined.

[00122] **Figure 80** shows deaminase protein lipofection to HEK cells using a GGCACTGCGGCTGGAGGTGGGGG (SEQ ID NO: 708) spacer. The HEK-4 on target, off target site 1, site 3, and site 4.

[00123] **Figure 81** shows the results of an *in vitro* assay for sgRNA activity for sgHR_13 (GTCAGGTCGAGGGTTCTGTC (SEQ ID NO: 709) spacer; C8 target: G51 to STOP), sgHR_14 (GGGCCGCAGTATCCTCACTC (SEQ ID NO: 710) spacer; C7 target; C7 target: Q68 to STOP), and sgHR_15 (CCGCCAGTCCCAGTACGGGA (SEQ ID NO: 711) spacer; CIO and C I 1 are targets: W239 or W237 to STOP).

[00124] **Figure 82** shows the results of an *in vitro* assay for sgHR_17 (CAACCACTGCTCAAAGATGC (SEQ ID NO: 712) spacer; C4 and C5 are targets: W410 to STOP), and sgHR_16 (CTTCCAGGATGAGAACACAG (SEQ ID NO: 713) spacer; C4 and C5 are targets: W273 to STOP).

[00125] **Figure 83** shows the direct injection of BE3 protein complexed with sgHR_13 in zebrafish embryos.

[00126] **Figure 84** shows the direct injection of BE3 protein complexed with sgHR_16 in zebrafish embryos.

[00127] **Figure 85** shows the direct injection of BE3 protein complexed with sgHR_17 in zebrafish embryos.

[00128] **Figure 86** shows exemplary nucleic acid changes that may be made using base editors that are capable of making a cytosine to thymine change.

[00129] **Figure 87** shows an illustration of apolipoprotein E (APOE) isoforms, demonstrating how a base editor *(e.g.,* BE3) may be used to edit one APOE isoform *(e.g.,* APOE4) into another APOE isoform *(e.g.,* APOE3r) that is associated with a decreased risk of Alzheimer's disease.

[00130] **Figure 88** shows base editing of APOE4 to APOE3r in mouse astrocytes. This figure depicts SEQ ID Nos: 671 and 627.

[00131] **Figure 89** shows base editing of PRNP to cause early truncation of the protein at arginine residue 37. This figure depicts SEQ ID Nos: 577 and 714.

[00132] **Figure 90** shows that knocking out UDG (which UGI inhibits) dramatically improves the cleanliness of efficiency of C to T base editing.

[00133] **Figure 91** shows that use of a base editor with the nickase but without UGI leads to a mixture of outcomes, with very high indel rates.

[00134] **Figures 92A** to **92G** show that SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, and VRER-BE3 mediate efficient base editing at target sites containing non-NGG PAMs in human cells. Figure 92A shows base editor architectures using *S. pyogenes* and *S. aureus* Cas9. Figure 92B shows recently characterized Cas9 variants with alternate or relaxed PAM requirements. Figures 92C and 92D show HEK293T cells treated with the base editor variants shown as described in Example 12. The percentage of total DNA sequencing reads (with no enrichment for transfected cells) with C converted to T at the target positions indicated are shown. The PAM sequence of each target tested is shown below the X-axis. The charts show the results for SaBE3 and SaKKH-BE3 at genomic loci with NNGRRT PAMs (Figure 92C), SaBE3 and SaKKH-BE3 at genomic loci with NNNRRT PAMs (Figure 92D), VQR-BE3 and EQR-BE3 at genomic loci with NGAG PAMs (Figure 92E), and with NGAH PAMs (Figure 92F), and VRER-BE3 at genomic loci with NGCG PAMs (Figure 92G). Values and error bars reflect the mean and standard deviation of at least two biological replicates.

[00135] **Figures 93A** to **93C** demonstrate that base editors with mutations in the cytidine deaminase domain exhibit narrowed editing windows. Figures 93A to 93C show HEK293T cells transfected with plasmids expressing mutant base editors and an appropriate sgRNA. Three days after transfection, genomic DNA was extracted and analyzed by high-throughput DNA sequencing at the indicated loci. The percentage of total DNA sequencing reads (without

enrichment for transfected cells) with C changed to T at the target positions indicated are shown for the EMXl site (SEQ ID NO: 721), HEK293 site 3 (SEQ ID NO: 719), FANCF site (SEQ ID NO: 722), HEK293 site 2 (SEQ ID NO: 720), site A (SEQ ID NO: 715), and site B (SEQ ID NO: 718) loci. Figure 93A illustrates certain cytidine deaminase mutations which narrow the base editing window. See Figure 98 for the characterization of additional mutations. Figure 93B shows the effect of cytidine deaminase mutations which effect the editing window width on genomic loci. Combining beneficial mutations has an additive effect on narrowing the editing window. Figure 93C shows that YE1-BE3, YE2-BE3, EE-BE3, and YEE-BE3 effect the product distribution of base editing, producing predominantly singly-modified products in contrast with BE3. Values and error bars reflect the mean and standard deviation of at least two biological replicates.

[00136] **Figures 94A** and **94B** show genetic variants from ClinVar that in principle can be corrected by the base editors developed in this work. The NCBI ClinVar database of human genetic variations and their corresponding phenotypes was searched for genetic diseases that in theory can be corrected by base editing. Figure 94A demonstrates improvement in base editing targeting scope among all pathogenic T→C mutations in the ClinVar database through the use of base editors with altered PAM specificities. The white fractions denote the proportion of pathogenic T→C mutations accessible on the basis of the PAM requirements of either BE3, or BE3 together with the five modified-PAM base editors developed in this work. Figure 94B shows improvement in base editing targeting scope among all pathogenic T→C mutations in the ClinVar database through the use of base editors with narrowed activity windows. BE3 was assumed to edit Cs in positions 4-8 with comparable efficiency as shown in Figures 93A to 93C. YEE-BE3 was assumed to edit with C5>C6>C7>others preference within its activity window. The white fractions denote the proportion of pathogenic T→C mutations that can be edited BE3 without comparable editing of other Cs (left), or that can be edited BE3 or YEE-BE3 without comparable editing of other Cs (right).

[00137] **Figures 95A** to **95B** show the effect of truncated guide RNAs on base editing window width. HEK293T cells were transfected with plasmids expressing BE3 and sgRNAs of different 5' truncation lengths. The treated cells were analyzed as described in the Examples. **Figure 95A** shows protospacer and PAM sequence (top, SEQ ID NO: 715) and cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with Ts at the target positions indicated, at a site within the EMXl genomic locus. At this site, the base editing window was altered through the use of a 17-nt truncated gRNA. **Figure 95B** shows protospacer and PAM sequences (top, SEQ ID NOs: 715 and 716) and

cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with Ts at the target positions indicated, at sites within the HEK site 3 and site 4 genomic loci. At these sites, no change in the base editing window was observed, but a linear decrease in editing efficiency for all substrate bases as the sgRNA is truncated was noted.

[00138] **Figure 96** shows the effect of APOBECl-Cas9 linker lengths on base editing window width. HEK293T cells were transfected with plasmids expressing base editors with rAPOBECl-Cas9 linkers of XTEN, GGS, $(GGS)_3$ (SEQ ID NO: 610), $(GGS)_5$ (SEQ ID NO: 610), or $(GGS)_7$ (SEQ ID NO: 610) and an sgRNA. The treated cells were analyzed as described in the Examples. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with Ts at the target positions indicated, are shown for the various base editors with different linkers.

[00139] **Figures 97A** to **97C** show the effect of rAPOBEC mutations on base editing window width. **Figure 97C** shows HEK293T cells transfected with plasmids expressing an sgRNA targeting either Site A or Site B and the BE3 point mutants indicated. The treated cells were analyzed as described in the Examples. All C's in the protospacer and within three basepairs of the protospacer are displayed and the cellular C to T conversion percentages are shown. The 'editing window widths', defined as the calculated number of nucleotides within which editing efficiency exceeds the half-maximal value, are displayed for all tested mutants.

[00140] **Figure 98** shows the effect of APOBEC1 mutation son product distributions of base editing in mammalian cells. HEK293T cells were transfected with plasmids expressing BE3 or its mutants and an appropriate sgRNAs. The treated cells were analyzed as described in the Examples. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with Ts at the target positions indicated, are shown (left). Percent of total sequencing reads containing the C to T conversion is shown on the right. The BE3 point mutants do not significantly affect base editing efficiencies at HEK site 4, a site with only one target cytidine.

[00141] **Figure 99** shows a comparison of on-target editing plasma delivery in BE3 and HF-BE3.

[00142] **Figure 100** shows a comparison of on-target editing in protein and plasma delivery of BE3.

[00143] **Figure 101** shows a comparison of on-target editing in protein and plasma devliery of HF-BE3.

[00144] **Figure 102** shows that both lipofection and installing HF mutations decrease off-target deamination events. The diamond indicates no off targets were detected and the specificity ratio was set to 100.

[00145] **Figure 103** shows *in vitro* C to T editing on a synthetic substrate with Cs placed at even positions in the protospacer (NNNNTC$_2$TC$_4$TC$_6$TC$_8$TCioTCi $_2$TCi $_4$TCi $_6$TCi $_8$TC$_2$oNGG, SEQ ID NO: 723).

[00146] **Figure 104** shows *in vitro* C to T editing on a synthetic substrate with Cs placed at odd positions in the protospacer (NNNNTC$_2$TC$_4$TC$_6$TC$_8$TCioTCi $_2$TCi$_4$TCi $_6$TCi $_8$TC$_2$oNGG, SEQ ID NO: 723).

[00147] **Figure 105** includes two graphs depicting the specificity ratio of base editing with plasmid vs. protein delivery.

[00148] **Figures 106A** to **106B** shows BE3 activity on non-NGG PAM sites. HEK293T cells were transfected with plasmids expressing BE3 and appropriate sgRNA. The treated cells were analyzed as described in the Examples. **Figure 106A** shows BE3 activity on sites can be efficiently targeted by SaBE3 or SaKKH-BE3. BE3 shows low but significant activity on the NAG PAM. This figure depicts SEQ ID NOs: 728 and 729. **Figure 106B** shows BE3 has significantly reduced editing at sites with NGA or NGCG PAMs, in contrast to VQR-BE3 or VRER-BE3. This figure depicts SEQ ID NOs: 730 and 731.

[00149] **Figures 107A** to **107B** show the effect of APOBEC1 mutations on VQR-BE3 and SaKKH-BE3. HEK293T cells were transfected with plasmids expressing VQR-BE3, SaKKH-BE3 or its mutants and an appropriate sgRNAs. The treated cells were analyzed as described in the Examples below. Cellular C to T conversion percentages, defined as the percentage of total DNA sequencing reads with Ts at the target positions indicated, are shown. **Figure 107A** shows that the window-modulating mutations can be applied to VQR-BE3 to enable selective base editing at sites targetable by NGA PAM. This figure depicts SEQ ID NOs: 732 and 733. **Figure 107B** shows that, when applied to SaKKH-BE3, the mutations cause overall decrease in base editing efficiency without conferring base selectivity within the target window. This figure depicts SEQ ID NOs: 728 and 734.

[00150] **Figure 108** shows a schematic representation of nucleotide editing. The following abbreviations are used: (MMR) - mismatch repair, (BE3 Nickase) - refers to base editor 3, which comprises a Cas9 nickase domain, (UGI) - uracil glycosylase inhibitor, (UDG) - uracil DNA glycosylase, (APOBEC) - refers to an APOBEC cytidine deaminase.

[00151] **Figure 109** shows schematic representations of exemplary base editing constructs. The structural arrangement of base editing constructs is shown for BE3, BE4-pmCDAl, BE4-

hAID, BE4-3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE4-XTEN, BE4-32aa, BE4-2xUGI, and BE4. Linkers are shown in grey (XTEN, SGGS (SEQ ID NO: 606), $(GGS)_3$ (SEQ ID NO: 610), and 32aa). Deaminases are shown (rAPOBECl, pmCDAl, hAID, and hAPOBEC3G). Uracil DNA Glycosylase Inhibitor (UGI) is shown. Single-stranded DNA binding protein (SSB) is shown in purple. Cas9 nickase, dCas9(A840H), is shown in red. Figure 109 also shows the following target sequences: EMXl, FANCF, HEK2, HEK3, HEK4, and RNF2. The amino acid sequences are indicated in SEQ ID NOs: 127-132 from top to bottom. The PAM sequences are the last three nucleotides. The target cytosine (C) is numbered and indicated in red.

[00152] **Figure 110** shows the base editing results for the indicated base editing constructs (BE3, pmCDAl hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, and BE4-2xUGI) on the targeted cytoine (C_5) of the EMXl sequence, GAGTCsCGAGCAGAAGAAGAAGGG (SEQ ID NO: 127). The total percentage of targeted cytosines (C_5) that were mutated is indicated for each base editing construct, under "C5". The total percentage of indels is indicated for each base editing construct, under "indel". The proportion of mutated cytosines that were mutated to an adenine (A), guanine (G), or thymine (T) are indicated for each base editing construct in the pie chart.

[00153] **Figure 111** shows the base editing results for the indicated base editing constructs (BE3, pmCDAl hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, and BE4-2xUGI) on the targeted cytoine (C_8) of the FANCF sequence, GGAATCCCsTTCTGCAGCACCTGG (SEQ ID NO: 128). The total percentage of targeted cytosines (C_8) that were mutated are indicated for each base editing construct, under "C_8". The total percentage of indels are indicated for each base editing construct, under "indel". The proportion of mutated cytosines that were mutated to an adenine (A), guanine (G), or thymine (T) are indicated for each base editing construct in the pie chart.

[00154] **Figure 112** shows the base editing results for the indicated base editing constructs (BE3, pmCDAl hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, and BE4-2xUGI) on the targeted cytoine (C_6) of the HEK2 sequence, GAACAC$_6$AAAGCATAGACTGCGGG (SEQ ID NO: 129). The total percentage of targeted cytosines (C_6) that were mutated are indicated for each base editing construct, under "C_6". The total percentage of indels are indicated for each base editing construct, under "indel". The proportion of mutated cytosines that were mutated to an adenine (A), guanine (G), or thymine (T) are indicated for each base editing construct in the pie chart.

[00155] **Figure 113** shows the base editing results for the indicated base editing constructs (BE3, pmCDAl hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, and BE4-2xUGI) on the targeted cytoine (C_5) of the HEK3 sequence, GGCCCsAGACTGAGCACGTGATGG (SEQ ID NO: 130). The total percentage of targeted cytosines (C_5) that were mutated are indicated for each base editing construct, under "C5.". The total percentage of indels are indicated for each base editing construct, under "indel". The proportion of mutated cytosines that were mutated to an adenine (A), guanine (G), or thymine (T) are indicated for each base editing construct in the pie chart.

[00156] **Figure 114** shows the base editing results for the indicated base editing constructs (BE3, pmCDAl hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, and BE4-2xUGI) on the targeted cytoine (C_5) of the HEK4 sequence, GGCAC_5TGCGGCTGGAGGTCCGGG (SEQ ID NO: 131). The total percentage of targeted cytosines (C_5) that were mutated are indicated for each base editing construct, under "C5.". The total percentage of indels are indicated for each base editing construct, under "indel". The proportion of mutated cytosines that were mutated to an adenine (A), guanine (G), or thymine (T) are indicated for each base editing construct in the pie chart.

[00157] **Figure 115** shows the base editing results for the indicated base editing constructs (BE3, pmCDAl hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, and BE4-2xUGI) on the targeted cytoine (C_6) of the RNF2 sequence, GTCATC$_6$TTAGTCATTACCTGAGG (SEQ ID NO: 132). The total percentage of targeted cytosines (C_6) that were mutated are indicated for each base editing construct, under "C_6.". The total percentage of indels are indicated for each base editing construct, under "indel". The proportion of mutated cytosines that were mutated to an adenine (A), guanine (G), or thymine (T) are indicated for each base editing construct in the pie chart.

[00158] **Figure 116** shows exemplary fluorescent labeled (Cy3 labeled) DNA constructs used to test for Cpfl mutants that nick the target strand. In the DNA construct of 1, both the non-target strand (top strand) and target strand (bottom strand) are fluorescently labeled. In the DNA construct of 2, the non-target strand (top strand) is fluorescently labeled and the target strand (bottom strand) is not fluorescently labeled. In the DNA construct of 3, the non-target strand (top strand) is not fluorescently labeled and the target strand (bottom strand) is fluorescently labeled.

[00159] **Figure 117** shows data demonstrating the ability of various Cpfl constructs *(e.g.,* R836A, R1138A, wild-type) to cleave the target and non-target strands of the DNA constructs

shown in Figure 116 over the reaction time of either 30 minutes (30 min) or greater than two hours (2h+).

[00160] **Figure 118** shows data demonstrating that a base editor having the architecture, APOBEC-AsCpfl(R912A)-UGI is capable of editing C residues *(e.g.,* of target sequences FANCF1, FANCF2, HEK3-3, and HEK3-4) having a window from the 7th to the 11th base of the target sequence. BG indicates background mutation levels (untreated). AsCpfl indicates AsCpfl only treated (control), APOBEC-AsCpfl(R912A)-UGI indicates a base editor containing a Cpfl that preferentially cuts the target strand, and APOBEC-AsCpfl(R1225A)-UGI indicates a self-defeating base editor containing a Cpfl that cuts the non-target strand. The target sequences of FANCF1, FANCF2, HEK3-3, and HEK3-4 are as follows:

FANCF1 GCGGATGTTCCAATCAGTACGCA (SEQ ID NO: 724)

FANCF2 CGAGCTTCTGGCGGTCTCAAGCA (SEQ ID NO: 725)

HEK3-3 TGCTTCTCCAGCCCTGGCCTGG (SEQ ID NO: 726)

HEK3-4 AGACTGAGCACGTGATGGCAGAG (SEQ ID NO: 727)

[00161] **Figure 119** shows a schematic representation of a base editor comprising a Cpfl protein *(e.g.,* AsCpfl or LbCpfl). Different linker sequences *(e.g.,* XTEN, GGS, $(GGS)_3$ (SEQ ID NO: 610), $(GGS)_5$ (SEQ ID NO: 610), and $(GGS)_7$ (SEQ ID NO: 610)) were tested for the portion labeled "linker," results of which are shown in Figure 120.

[00162] **Figure 120** shows data demonstrating the ability of the construct shown in Figure 119 to edit the C_8 residue of the HEK3 site TGCTTCTC$_8$CAGCCCTGGCCTGG (SEQ ID NO: 592). Different linker sequences, which link the APOBEC domain to the Cpfl domain *(e.g.,* LbCpfl(R836A) or AsCpfl (R912A)) were tested. Exemplary linkers that were tested include XTEN, GGS, $(GGS)_3$(SEQ ID NO: 610), $(GGS)_5$(SEQ ID NO: 610), and $(GGS)_7$(SEQ ID NO: 610).

[00163] **Figure 121** shows data demonstrating the ability of the construct shown in Figure 119, having the LbCpfl domain, to edit the C_8 and C_9 residues of the HEK3 TGCTTCTC$_8$C$_9$AGCCCTGGCCTGG (SEQ ID NO: 592). Different linker sequences from a database maintained by the Centre of Integrative Bioinformatics VU, which link the APOBEC domain to the LbCpfl domain were tested. Exemplary linkers that were tested include lau7, lclk, lc20, lee8, lflz, lign, ljmc, lsfe, 2ezx, and 2reb.

[00164] **Figure 122** shows a schematic representation of the structure of AsCpfl, where the N and C termini are indicated.

[00165] **Figure 123** shows a schematic representation of the structure of SpCas9, where the N and C termini are indicated.

[00166] **Figure 124** shows a schematic representation of AsCpfl, where the red circle indicates the predicted area where the editing window is. The square indicates a helical region that may be obstructing APOBEC activity.

[00167] **Figures 125A** and **125B** show engineering and *in vitro* characterization of a high fidelity base editor (HF-BE3). **Figure 125A** shows a schematic representation of HF-BE3. Point mutations introduced into BE3 to generate HF-BE3 are shown. The representation used PDB structures 4UN3 (Cas9), 4ROV (cytidine deaminase) and 1UGI (uracil DNA glycosylase inhibitor). **Figure 125B** shows *in vitro* deamination of synthetic substrates containing 'TC repeat protospacers. Values and error bars reflect mean and range of two independent replicates performed on different days.

[00168] **Figures 126A** to **126C** show purification of base editor proteins. **Figure 126A** shows selection of optimal *E. coli* strain for base editor expression. After IPTG-induced protein expression for 16 h at 18 °C, crude cell lysate was analyzed for protein content. BL21 Star (DE3) (Thermo Fisher) cells showed the most promising post-expression levels of both BE3 and HF-BE3 and were used for expression of base editors. **Figure 126B** shows purification of expressed base editor proteins. Placing the His6 tag on the C-terminus of the base editors lead to production of a truncation product for both BE3 and HF-BE3 (lanes 1 and 2). Unexpectedly, this truncation product was removed by placing the His6 tag on the N-terminus of the protein (lanes 3-6). Inducing expression of base editors at a cell density of OD600 = 0.7 (lanes 4-5), later than is optimal for Cas9 expression (OD600 = 0.4)1, improves yield of base editor proteins. Purification was performed using a manual HisPur resin column followed by cation exchange FPLC (Akta). **Figure 126C** shows purified BE3 and HF-BE3. Different concentrations of purified BE3 and HF-BE3 were denatured using heat and LDS and loaded onto a polyacrylamide gel. Protein samples are representative of proteins used in this study. Gels in **Figures 126A** to **126C** are BOLT Bis-Tris Plus 4-12% polyacrylamide (Thermo Fisher). Electrophoresis and staining were performed as described in Methods.

[00169] **Figures 127A** to **127D** show activity of a high fidelity base editor (HF-BE3) in human cells. **Figures 127A** to **127C** show on- and off-target editing associated with plasmid transfection of BE3 and HF-BE3 was assayed using high-throughput sequencing of genomic DNA from HEK293T cells treated with sgRNAs targeting non-repetitive genomic loci EMX1 **(Figure 127A),** FANCF **(Figure 127B),** and HEK293 site 3 **(Figure 127C).** On- and off-target loci associated with each sgRNA are separated by a vertical line. **Figure 127D** shows on- and off-target editing associated with the highly repetitive sgRNA targeting VEGFA site 2. Values and error bars reflect mean ± S.D. of three independent biological replicates performed on

different days. For **Figures 127A** to **127C,** stars indicate significant editing based on a comparison between the treated sample and an untreated control. * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$ (Student's two tailed t-test). For **Figure 127D,** asterisks are not shown since all treated samples displayed significant editing relative to the control. Individual p-values are listed in in Table 16.

[00170] **Figures 128A** to **128C** show the effect of dosage of BE3 protein or plasmid on the efficiency of on-target and off-target base editing in cultured human cells. **Figure 128A** shows on-target editing efficiency at each of the four genomic loci was averaged across all edited cytosines in the activity window for each sgRNA. Values and error bars reflect mean ± S.E.M of three independent biological replicates performed on different days. **Figures 128B** and **128C** show on- and off-target editing at the EMX1 site arising from BE3 plasmid titration **(Figure 128B)** or BE3 protein titration **(Figure 128C)** in HEK293T cells. Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days.

[00171] **Figures 129A** to **129B** show on-target:off-target base editing frequency ratios for plasmid and protein delivery of BE3 and HF-BE3. Base editing on-target:off-target specificity ratios were calculated by dividing the on-target editing percentage at a particular cytosine in the activity window by the off-target editing percentage at the corresponding cytosine for the indicated off-target locus (see Methods). When off-target editing was below the threshold of detection (0.025% of sequencing reads), we set the off-target editing to the limit of detection (0.025%) and divided the on-target editing percentage by this upper limit. In these cases, denoted by ♦, the specificity ratios shown represent lower limits. Specificity ratios are shown for non-repetitive sgRNAs FANCF, HEK 293 site 3, and FANCF **(Figure 129A)** and for the highly repetitive sgRNA VEGFA site 2 **(Figure 129B).** Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days.

[00172] **Figures 130A** to **130D** show protein delivery of base editors into cultured human cells. **Figures 130A** to **130D** show on- and off-target editing associated with RNP delivery of base editors complexed with sgRNAs targeting EMX1 **(Figure 130A),** FANCF **(Figure 130B),** HEK293 site 3 **(Figure 130C)** and VEGFA site 2 **(Figure 130D).** Off-target base editing was undetectable at all of the sequenced loci for non-repetitive sgRNAs. Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days. Stars indicate significant editing based on a comparison between the treated sample and an untreated control. * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$ (Student's two tailed t-test).

[00173] **Figures 131A** to **131C** show indel formation associated with base editing at genomic loci. **Figure 131A** shows indel frequency at on-target loci for VEGFA site 2, EMXl, FANCF, and HEK293 site 3 sgRNAs. **Figure 131B** shows the ratio of base editingdndel formation. The diamond (♦) indicates no indels were detected (no significant difference in indel frequency in the treated sample and in the untreated control). **Figure 131C** shows indels observed at the off-target loci associated with the on-target sites interrogated in **Figure 131A.** Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days.

[00174] **Figures 132A** to **132D** show DNA-free *in vivo* base editing in zebrafish embryos and in the inner ear of live mice using RNP delivery of BE3. **Figure 132A** shows on-target genome editing in zebrafish harvested 4 days after injection of BE3 complexed with indicated sgRNA. Values and error bars reflect mean ± s.d. of three injected and three control zebrafish. Controls were injected with BE3 complexed with an unrelated sgRNA. **Figure 132B** shows schematic showing *in vivo* injection of BE3:sgRNA complexes encapsulated into cationic lipid nanoparticles **Figure 132C** shows base editing of cytosine residues in the base editor window at the VEGFA site 2 genomic locus. **Figure 132D** shows on-target editing at each cytosine in the base editing window of the VEGFA site 2 target locus. **Figure 132D (Figures 132C** and **132D)** shows values and error bars reflect mean ± S.E.M. of three mice injected with sgRNA targeting VEGFA Site 2, three uninjected mice and one mouse injected with unrelated sgRNA.

[00175] **Figures 133A** to **133E** show on- and off-target base editing in murine NIH/3T3 cells. **Figure 133A** shows on-target base editing associated with the 'VEGFA site 2' sgRNA (See Figure **132E** for sequences). The negative control corresponds to cells treated with plasmid encoding BE3 but no sgRNA. Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days. **Figures 133B** to **133E** show off-target editing associated with this site was measured using high-throughput DNA sequencing at the top four predicted off-target loci for this sgRNA (sequences shown in **Figure 132E).** **Figure 133B** shows off-target 2, **Figure 133C** shows off-target 1, **Figure 133D** shows off-target 3, **Figure 133E** shows off-target 4. Values and error bars reflect mean ± S.D. of three independent biological replicates performed on different days.

[00176] **Figures 134A** to **134B** show off-target base editing and on-target indel analysis from *in* vzvo-edited murine tissue. **Figure 134A** shows editing plotted for each cytosine in the base editing window of off-target loci associated with VEGFA site 2. **Figure 134B** shows indel rates at the on-target base editor locus. Values and error bars reflect mean ± S.E.M of three injected and three control mice.

[00177] **Figures** 135A to 135C show the effects on base editing product purity of knocking out UNG. **Figure** 135A shows HAP1 (UNG^+) and HAP1 UNG^- cells treated with BE3 as described in the Materials and Methods of Example **17.** The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. **Figure** 135B shows protospacers and PAM sequences of the genomic loci tested, with the target Cs analyzed in **Figure** 135A shown in red. **Figure** 135C shows the frequency of indel formation following treatment with BE3 in HAP1 cells or HAP1 UNG^- cells. Values and error bars reflect the mean ± S.D. of three independent biological replicates performed on different days.

[00178] **Figures** 136A to 136D show the effects of multi-C base editing on product purity. **Figure** 136A shows representative high-throughput sequencing data of untreated, BE3-treated, and AID-BE3-treated human HEK293T cells. The sequence of the protospacer is shown at the top, with the PAM and the target Cs in red with subscripted numbers indicating their position within the protospacer. Underneath each sequence are the percentages of total sequencing reads with the corresponding base. The relative percentage of target Cs that are cleanly edited to T rather than to non-T bases is much higher for AID-BE3-treated cells, which edits three Cs at this locus, than for BE3-treated cells, which edits only one C. **Figure** 136B shows HEK293T cells treated with BE3, CDA1-BE3, and AID-BE3 as described in the Materials and Methods of Example **17.** The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. **Figure** 136C shows protospacers and PAM sequences of genomic loci studied, with the target Cs that are analyzed in **Figure** 136B shown in red. **Figure** 136D shows the frequency of indel formation following the treatment shown in **Figure** 136A. Values and error bars reflect the mean ± S.D. of three independent biological replicates performed on different days.

[00179] **Figures** 137A to 137C show the effects on C-to-T editing efficiencies and product purities of changing the architecture of BE3. **Figure** 137Ashows protospacers and PAM sequences of genomic loci studied, with the target Cs in **Figure** 137C shown in purple and red, and the target Cs in **Figure** 137B shown in red. **Figure** 137Bshows HEK293T cells treated with BE3, SSB-BE3, N-UGI-BE3, and BE3-2xUGI as described in the Materials and Methods of Example **17.** The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown for BE3, N-UGI-BE3, and BE3-2xUGI. **Figure** 137Cshows C-to-T base editing efficiencies. Values and error bars reflect the mean ± S.D. of three independent biological replicates performed on different days.

[00180] **Figures** 138A to 138D show the effects of linker length variation in BE3 on C-to-T editing efficiencies and product purities. **Figure** 138A shows the architecture of BE3, BE3C,

BE3D, and BE3E **Figure 138B** shows protospacers and PAM sequences of genomic loci studied, with the target Cs in **Figure 138C** shown in purple and red, and target Cs in **Figure 138D** shown in red. **Figure 138C** shows HEK293T cells treated with BE3, BE3C, BE3D, or BE3E as described in the Materials and Methods of Example 17. C-to-T base editing efficiencies are shown. **Figure 138D** shows the product distribution among edited DNA sequencing reads (reads in which the target C is mutated) for BE3, BE3C, BE3D, and BE3E. Values and error bars reflect the mean ± S.D. of three independent biological replicates performed on different days.

[00181] **Figures 139A** to **139D** show BE4 increases base editing efficiency and product purities compared to BE3. **Figure 139A** shows the architectures of BE3, BE4, and Target-AID. **Figure 139B** shows protospacers and PAM sequences of genomic loci studied, with the target Cs in **Figure 139C** shown in purple and red, and the target Cs in **Figure 139D** shown in red. **Figure 139C** shows HEK293T cells treated with BE3, BE4, or Target-AID as described in the Materials and Methods of Example 17. C-to-T base editing efficiencies are shown. **Figure 139D** shows the product distribution among edited DNA sequencing reads (reads in which the target C is mutated) for BE3 and BE4. Values and error bars reflect the mean ± S.D. of three independent biological replicates performed on different days.

[00182] **Figures 140A** to **140C** show CDA1-BE3 and AID-BE3 edit Cs following target Gs more efficiently than BE3. **Figure 140A** shows protospacer and PAM sequences of genomic loci studied, with target Cs edited by BE3, CDA1-BE3, and AID-BE3 shown in red, and target Cs (following Gs) edited by CDA1-BE3 and AID-BE3 only shown in purple. **Figure 140B** shows HEK293T cells treated with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 as described in the Materials and Methods of Example 17. C-to-T base editing efficiencies are shown. **Figure 140C** shows individual DNA sequencing reads from HEK293T cells that were treated with BE3, CDA1-BE3, or AID-BE3 targeting the *HEK2* locus and binned according to the sequence of the protospacer and analyzed, revealing that > 85% of sequencing reads that have clean C to Tedits by CDA1-BE3 and AID-BE3 have both Cs edited to T **(Figure 140C).**

[00183] **Figures 141A** to **141C** show uneven editing in sites with multiple editable Cs results in lower product purity. **Figure 141A** shows protospacers and PAM sequences of genomic loci studied, with the target Cs in **Figure 141C** shown in purple and red, and target Cs in **Figure 141B** shown in red. **Figures 141B** and **141C** show HEK293T cells treated with BE3 as described in the Materials and Methods of Example 17. The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. C to non-T editing is more frequent when editing efficiencies are unequal for two Cs within the

same locus. Values and error bars reflect the mean ± S.D. of three independent biological replicates performed on different days.

[00184] **Figures 142A** to **142D** show base editing of multiple Cs results in higher base editing product purity. **Figure 142A** shows protospacers and PAM sequences of genomic loci studied, with the target Cs that are investigated in **Figure 142B** shown in red. **Figure 142B** shows HEK293T cells treated with BE3 or BE3B (which lacks UGI) as described in the Materials and Methods of Example 17. The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. **Figure 142C** shows the HTS reads from HEK293T cells that were treated with BE3 or BE3B (which lacks UGI) targeting the *HEK2* locus were binned according to the identity of the primary target C at position 6. The resulting reads were then analyzed for the identity of the base at the secondary target C at position 4. C_6 is more likely to be incorrectly edited to a non-T when there is only a single editing event in that read. **Figure 142D** shows the distribution of edited reads with A, G, and T at C_5 in cells treated with BE3 or BE3B targeting the *HEK4* locus (a site with only a single editable C), illustrating that single G:U mismatches are processed via UNG-initiated base excision repair to give a mixture of products. Values and error bars reflect the mean ± S.D. of three independent biological replicates performed on different days.

[00185] **Figure 143** shows base editing of multiple Cs results in higher base editing product purity at the *HEK3* and *RNF2* loci. DNA sequencing reads from HEK293T cells treated with BE3 or BE3B (without UGI) targeting the *HEK3* and *RNF2* loci were separated according to the identity of the base at the primary target C position (in red). The four groups of sequencing reads were then interrogated for the identity of the base at the secondary target C position (in purple). For BE3, when the primary target C (in red) is incorrectly edited to G, the secondary target C is more likely to remain C. Conversely, when the primary target C (in red) is converted to T, the secondary target C is more likely to also be edited to a T in the same sequencing read. Values and error bars reflect the mean ± S.D. of three independent biological replicates performed on different days.

[00186] **Figures 144A** to **144C** show BE4 induces lower indel frequencies than BE3, and Target-AID exhibits similar product purities as CDA1-BE3. **Figure 144A** shows HEK293T cells treated with BE3, BE4, or Target-AID as described in the Materials and Methods of Example 17. The frequency of indel formation (see Materials and Methods of Example 17) is shown. **Figure 144B shows** HEK293T cells treated with CDA1-BE3 or Target-AID as described in the Materials and Methods of Example 17. The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. **Figure 144C** shows

protospacers and PAM sequences of genomic loci studied, with the target Cs that are investigated in **Figure 144B** shown in red. Values and error bars reflect the mean ± S.D. of three independent biological replicates performed on different days.

[00187] **Figures 145A** to **145C** show SaBE4 exhibits increased base editing yields and product purities compared to SaBE3. **Figure 145A** shows HEK293T cells treated with SaBE3 and SaBE4 as described in the Materials and Methods of Example 17. The percentage of total DNA sequencing reads with Ts at the target positions indicated are shown. **Figure 145B** shows protospacers and PAM sequences of genomic loci studied, with the target Cs in **Figure 145A** shown in purple and red, with target Cs that are investigated in **Figure 145C** shown in red. **Figure 145C** shows the product distribution among edited DNA sequencing reads (reads in which the target C is mutated). Values and error bars reflect the mean ± S.D. of three independent biological replicates performed on different days.

[00188] **Figure 146** shows base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *EMX1* locus. The sequence of the protospacer is shown at the top, with the PAM and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Materials and Methods of Example 17. Values shown are from one representative experiment.

[00189] **Figure 147** shows base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *FANCF* locus. The sequence of the protospacer is shown at the top, with the PAM and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Materials and Methods of Example 17. Values shown are from one representative experiment.

[00190] **Figure 148** shows base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *HEK2* locus. The sequence of the protospacer is shown at the top, with the PAM and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Materials and Methods of Example 17. Values shown are from one representative experiment.

[00191] **Figure 149** shows base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *HEK3* locus. The sequence of the protospacer is shown at the top, with the PAM and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total

sequencing reads with the corresponding base. Cells were treated as described in the Materials and Methods of Example 17. Values shown are from one representative experiment.

[00192] **Figure 150** shows base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *HEK4* locus. The sequence of the protospacer is shown at the top, with the PAM and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Materials and Methods of Example 17. Values shown are from one representative experiment.

[00193] **Figure 151** shows base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *RNF2* locus. The sequence of the protospacer is shown at the top, with the PAM and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Materials and Methods of Example 17. Values shown are from one representative experiment.

[00194] **Figure 152** shows a schematic of LBCpf 1 fusion constructs. Construct 10 has a domain arrangement of [Apobec]-[LbCpfl]-[UGI]-[UGI]; construct 11 has a domain arrangement of [Apobec]-[LbCpfl]-[UGI]; construct 12 has a domain arrangement of [UGI]-[Apobec]-[LbCpfl]; construct 13 has a domain arrangement of [Apobec]-[UGI]-[LbCpfl]; construct 14 has a domain arrangement of [LbCpfl]-[UGI]-[Apobec]; construct 15 has a domain arrangement of [LbCpfl]-[Apobec]-[UGI]. For each construct three different LbCpfl proteins were used (D/N/A, which refers to nuclease dead LbCpfl (D); LbCpfl nickase (N) and nuclease active LbCpfl (A)).

[00195] **Figure 153** shows the percentage of C to T editing of six C residues in the EMX target TTTGT A C$_3$TTTGTC9C $_1$oTC 12C$_{13}$GGTTC $_{18}$TG (SEQ ID NO: 738) using a guide of 19 nucleotides in length, *i.e.*, EMX19: TACTTTGTCCTCCGGTTCT (SEQ ID NO: 744). Editing was tested for several of the constructs shown in Figure 152.

[00196] **Figure 154** shows the percentage of C to T editing of six C residues in the EMX target TTTGT A C$_3$TTTGTC9C $_1$oTC 12C$_{13}$GGTTC $_{18}$TG (SEQ ID NO: 738) using a guide of 18 nucleotides in length, *i.e.*, EMX18: TACTTTGTCCTCCGGTTC (SEQ ID NO: 745). Editing was tested for several of the constructs shown in Figure 152.

[00197] **Figure 155** shows the percentage of C to T editing of six C residues in the EMX target TTTGT A C$_3$TTTGTC9C $_1$oTC 12C$_{13}$GGTTC $_{18}$TG (SEQ ID NO: 738) using a guide of 17 nucleotides in length, *i.e.*, EMX17: TACTTTGTCCTCCGGTT (SEQ ID NO: 746). Editing was tested for several of the constructs shown in Figure 152.

[00198] **Figure 156** shows the percentage of C to T editing of eight C residues in the HEK2 target TTTCCiAGC $_4$C$_5$C$_6$GC$_8$TGGCi $_2$Ci$_3$Ci$_4$TGTAAA (SEQ ID NO: 739) using a guide of 23 nucleotides in length, *i.e.,* Hek2_23: CAGCCCGCTGGCCCTGTAAAGGA (SEQ ID NO: 747). Editing was tested for several of the constructs shown in Figure 152.

[00199] **Figure 157** shows the percentage of C to T editing of eight C residues in the HEK2 target TTTCCiAGC $_4$C$_5$C$_6$GC$_8$TGGCi $_2$Ci$_3$Ci$_4$TGTAAA (SEQ ID NO: 739) using a guide of 20 nucleotides in length, *i.e.,* Hek2_20: CAGCCCGCTGGCCCTGTAAA (SEQ ID NO: 748). Editing was tested for several of the constructs shown in Figure 152.

[00200] **Figure 158** shows the percentage of C to T editing of eight C residues in the HEK2 target TTTCCiAGC $_4$C$_5$C$_6$GC$_8$TGGCi $_2$Ci$_3$Ci$_4$TGTAAA (SEQ ID NO: 739) using a guide of 19 nucleotides in length, *i.e.,* Hek2_19: CAGCCCGCTGGCCCTGTAA (SEQ ID NO: 749). Editing was tested for several of the constructs shown in Figure 152.

[00201] **Figure 159** shows the percentage of C to T editing of eight C residues in the HEK2 target TTTCCiAGC $_4$C$_5$C$_6$GC$_8$TGGCi $_2$Ci$_3$Ci$_4$TGTAAA (SEQ ID NO: 739) using a guide of 18 nucleotides in length, *i.e.,* Hek2_18: CAGCCCGCTGGCCCTGTA (SEQ ID NO: 750). Editing was tested for several of the constructs shown in Figure 152.

[00202] **Figure 160** shows the editing percentage values (after adjustment based on indel count), and the percentage of indels for the experiments depicted in figure 153.

[00203] **Figure 161** shows the editing percentage values (after adjustment based on indel count), and the percentage of indels for the experiments depicted in figure 154.

[00204] **Figure 162** shows the editing percentage values (after adjustment based on indel count), and the percentage of indels for the experiments depicted in figure 155.

[00205] **Figure 163** shows the editing percentage values (after adjustment based on indel count), and the percentage of indels for the experiments depicted in figure 156.

[00206] **Figure 164** shows the editing percentage values (after adjustment based on indel count), and the percentage of indels for the experiments depicted in figure 157.

[00207] **Figure 165** shows the editing percentage values (after adjustment based on indel count), and the percentage of indels for the experiments depicted in figure 158.

[00208] **Figure 166** shows the editing percentage values (after adjustment based on indel count), and the percentage of indels for the experiments depicted in figure 159.

DEFINITIONS

[00209] As used herein and in the claims, the singular forms "a," "an," and "the" include the singular and the plural reference unless the context clearly indicates otherwise. Thus, for example, a reference to "an agent" includes a single agent and a plurality of such agents.

[00210] The term "nucleic acid programmable DNA binding protein" or "napDNAbp" refers to a protein that associates with a nucleic acid *(e.g.,* DNA or RNA), such as a guide nucleic acid *(e.g.,* gRNA), that guides the napDNAbp to a specific nucleic acid sequence, for example, by hybridinzing to the target nucleic acid sequence. For example, a Cas9 protein can associate with a guide RNA that guides the Cas9 protein to a specific DNA sequence is has complementary to the guide RNA. In some embodiments, the napDNAbp is a class 2 microbial CRISPR-Cas effector. In some embodiments, the napDNAbp is a Cas9 domain, for example, a nuclease active Cas9, a Cas9 nickase (nCas9), or a nuclease inactive Cas9 (dCas9). Examples of nucleic acid programmable DNA binding proteins include, without limitation, Cas9 *(e.g.,* dCas9 and nCas9), CasX, CasY, Cpfl, C2cl, C2c2, C2C3, and Argonaute. It should be appreciated, however, that nucleic acid programmable DNA binding proteins also include nucleic acid programmable proteins that bind RNA. For example, the napDNAbp may be associated with a nucleic acid that guides the napDNAbp to an RNA. Other nucleic acid programmable DNA binding proteins are also within the scope of this disclosure, though they may not be specifically described in this disclosure.

[00211] In some embodiments, the napDNAby is an "RNA-programmable nuclease" or "RNA-guided nuclease." The terms are used interchangeably herein and refer to a nuclease that forms a complex with *(e.g.,* binds or associates with) one or more RNA(s) that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule. gRNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), though "gRNA" is also used to refer to guide RNAs that exist as either single molecules or as a complex of two or more molecules. Typically, gRNAs that exist as a single RNA species comprise two domains: (1) a domain that shares homology to a target nucleic acid *(i.e.,* directs binding of a Cas9 complex to the target); and (2) a domain that binds a Cas9 protein. In some embodiments, domain (2) corresponds to a sequence known as a tracrRNA and comprises a stem-loop structure. In some embodiments, domain (2) is identical or homologous to a tracrRNA as provided in Jinek *et al., Science* 337:816-821 (2012), the entire contents of which is incorporated herein by

reference. Other examples of gRNAs *(e.g.*, those including domain 2) can be found in U.S. Provisional Patent Application, U.S.S.N. 61/874,682, filed September 6, 2013, entitled "Switchable Cas9 Nucleases And Uses Thereof," and U.S. Provisional Patent Application, U.S.S.N. 61/874,746, filed September 6, 2013, entitled "Delivery System For Functional Nucleases," the entire contents of each are hereby incorporated by reference in their entirety. In some embodiments, a gRNA comprises two or more of domains (1) and (2), and may be referred to as an "extended gRNA." For example, an extended gRNA will bind two or more Cas9 proteins and bind a target nucleic acid at two or more distinct regions, as described herein. The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease/RNA complex to said target site, providing the sequence specificity of the nuclease:RNA complex. In some embodiments, the RNA-programmable nuclease is the (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (Csnl) from *Streptococcus pyogenes* (see, *e.g.*, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes."* Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C , Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607 (2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821 (2012), the entire contents of each of which are incorporated herein by reference.

[00212] In some embodiments, any of the sgRNAs provided herein comprise a sequence, e.g., a sgRNA backbone sequence that binds to a napDNAbp. For example sgRNAs have been described in Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, and Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 337, 816-812; Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. Nature Methods, 10, 957-963; Li JF, Norville JE, Aach J, McCromack M, Zhang D, Bush J, Church GM, and Sheen J (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nature Biotech, 31, 688-691; Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JRJ, Joung JK (2013) Efficient in vivo genome editing using RNA-guided nucleases. Nat Biotechnol, 31, 227-229; Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex

genome engineering using CRIPSR/Cas systems. Science, 339, 819-823; Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol, 31, 230-232; Jinek MJ, East A, Cheng A, Lin S, Ma E, Doudna J (2013) RNA-programmed genome editing in human cells. eLIFE, 2:e00471; DiCarlo JE, Norville JE, Mali P, Rios, Aach J, Church GM (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucl Acids Res, 41, 4336-4343; Briner AE, Donohoue PD, Gomaa AA, Selle K, Slorach EM, Nye CH, Haurwitz RE, Beisel CL, May AP, and Barrangou R (2014) Guide RNA functional modules direct Cas9 activity and orthogonality. Mol Cell, 56, 333-339; the contents of each of which are incorporated herein by reference. In some embodiments, any of the gRNAs *(e.g.,* sgRNAs) provided herin comprise the nucleic acid sequence of GTAATTTCTACTAAGTGTAGAT (SEQ ID NO: 741), wherein each of the Ts of SEQ ID NO: 741 are uracil (U), *i.e.,* GUAAUUUCUACUAAGUGUAGAU, or the sequence GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUUU-3' (SEQ ID NO: 618).

[00213] Because RNA-programmable nucleases *(e.g.,* Cas9) use RNA:DNA hybridization to target DNA cleavage sites, these proteins are able to target, in principle, any sequence specified by the guide RNA. Methods of using RNA-programmable nucleases, such as Cas9, for site-specific cleavage *(e.g.,* to modify a genome) are known in the art (see *e.g.,* Cong, L. *et ah,* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819-823 (2013); Mali, P. *et ah,* RNA-guided human genome engineering via Cas9. *Science* **339,** 823-826 (2013); Hwang, W.Y. *et ah,* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* **31,** 227-229 (2013); Jinek, M. *et ah* RNA-programmed genome editing in human cells. *eLife 2,* e00471 (2013); Dicarlo, J.E. *et ah,* Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research* (2013); Jiang, W. *et ah,* RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31,** 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

[00214] The term "Cas9" or "Cas9 nuclease" refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof *(e.g.,* a protein comprising an active, inactive, or partially active DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casnl nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable

elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gNRA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See, *e.g.*, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, *e.g.*, "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C , Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(201 1); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, a Cas9 nuclease has an inactive *{e.g.*, an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase.

[00215] A nuclease-inactivated Cas9 protein may interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9). Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (See, *e.g.,* Jinek *et al, Science.* 337:816-821(2012); Qi *et al,* "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell.* 28; 152(5): 1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvCl subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvCl subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al, Science.* 337:816-821(2012); Qi *et al, Cell.* 28; 152(5): 1173-83 (2013)). In some embodiments, proteins comprising fragments of Cas9 are provided. For example, in some embodiments, a protein comprises one of two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9. In some embodiments, proteins comprising Cas9 or fragments thereof are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to wild type Cas9. In some embodiments, the Cas9 variant may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more amino acid changes compared to wild type Cas9. In some embodiments, the Cas9 variant comprises a fragment of Cas9 *{e.g.,* a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas9. In some embodiments, the fragment is is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas9.

[00216] In some embodiments, the fragment is at least 100 amino acids in length. In some embodiments, the fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or at least 1300 amino acids in length. In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1, SEQ ID NO: 1 (nucleotide); SEQ ID NO: 2 (amino acid)).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGG
CGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAA
TACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGT
GGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACA
CGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGA
AAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGAC
AAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATC
ATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGA
TAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTA
TTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGA
TTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGA
ATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTTGATTTGG
CAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAA
TTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATT
TATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAG
GCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGAC
TCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT
TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGA
AGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAA
TTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACA
ACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGA
CAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCT
TGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTG
CATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGT
TGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATA
AAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTT
ACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAAC
CAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACA
AATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAAT
GTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGC
GCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAG
AAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGG
GGGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGA
TGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTG
ATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAAT
CAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACA
TTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATG
AACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACT
GTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCG
TTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGC

GAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTC
TTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTAT
TATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTT
TAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCA
ATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACG
TTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAA
CGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGA
GGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCG
CCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATAC
GATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAAT
TAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAAT
TACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAA
GAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATG
TTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATA
TTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATG
GAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGT
CTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAA
GTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCA
ATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATC
CAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTT
GCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTA
GGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAG
AAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATA
TAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAA
TTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTT
AGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACA
ATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGT
GAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGC
ATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCAT
TTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAAC
AATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCC
ATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGT
GACTGA (SEQ ID NO: 1)

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKR
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL

NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL
QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT
STKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:2)
(single underline: HNH domain; double underline: RuvC domain)

[00217] In some embodiments, wild type Cas9 corresponds to, or comprises SEQ ID NO:3 (nucleotide) and/or SEQ ID NO: 4 (amino acid):

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGC
TGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAAC
ACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGG
CGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACG
TCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAA
GTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAA
GAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCAT
GAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATA
AAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGG
CACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTT
CATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCA
AGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGC
TAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAA
CCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAG
CTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAA
TCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACC
TTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAG
GCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGA
CACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATT
CTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAA
GAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAG
AGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGA
CAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGA
AGGCAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAA
TCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGG
TTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGG
AAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTT
TGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAG
TATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGC
GTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATT
CAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAA
AATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGT
CACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGA
TAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTG
AAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACG
ATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTC
GCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGAT
TTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATG
ACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGA

CTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGC
ATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACA
AACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGG
GGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAA
CTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACG
AGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGA
ACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCT
TTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCG
AGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTA
TTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTA
ACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAAC
GTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTC
CCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGT
AATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATA
AAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGT
CGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTAT
GGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAG
ATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAA
GACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACC
AATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTG
AGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGA
CCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCAT
CGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACA
GTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAAC
TGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGA
AAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGA
TCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAA
CGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGT
CTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCA
CCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCG
ACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGC
CAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGT
GAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGC
CGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAG
GAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTC
GGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAG
TCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTA
CAAGGATGACGATGACAAGGCTGCAGGA (SEQ ID NO:3)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS

LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:4)
(single underline: HNH domain; double underline: RuvC domain)

[00218] In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_002737.2, SEQ ID NO: 5 (nucleotide); and Uniport Reference Sequence: Q99ZW2, SEQ ID NO: 6 (amino acid).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGG
CGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAA
TACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGT
GGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACA
CGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGA
AAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGAC
AAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATC
ATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGA
TAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTA
TTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGA
TTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGA
ATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTTGATTTGG
CAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAA
TTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATT
TATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAG
GCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGAC
TCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT
TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGA
AGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAA
TTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACA
ACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGA
CAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCT
TGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTG
CATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGT
TGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATA
AAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTT

ACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAAC
CAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACA
AATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAAT
GTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGT
ACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAG
AAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGG
GAGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGA
TGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTG
ATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAAT
CAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACA
TTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATG
AACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACT
GTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATA
TCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTC
GCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGAT
TCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCT
ATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCG
TTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTAAAGACGATT
CAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAA
CGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTA
AACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTG
GAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAAC
TCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAA
TACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTA
AATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAAC
AATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGAT
TAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATG
ATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAA
ATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAA
ATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAA
TTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCC
CCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGA
GTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGG
GATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGT
GGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTT
ACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTT
TTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTA
AATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGG
AGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTA
TATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAA
AACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAAT
CAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTA
GTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTA
TTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATA

CAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCT
TATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAG
GAGGTGACTGA (SEQ ID NO: 5)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 6) (single underline: HNH domain; double underline: RuvC domain)

[00219] In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquisl* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1), *Listeria innocua* (NCBI Ref: NP_472073.1), *Campylobacter jejuni* (NCBI Ref: YP_002344900.1) or *Neisseria. meningitidis* (NCBI Ref: YP_002342100.1) or to a Cas9 from any of the organisms listed in Example 5.

[00220] In some embodiments, dCas9 corresponds to, or comprises in part or in whole, a Cas9 amino acid sequence having one or more mutations that inactivate the Cas9 nuclease

activity. For example, in some embodiments, a dCas9 domain comprises D10A and/or H840A mutation.

dCas9 (D10A and H840A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH
ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD
LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEI
SGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT Y
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQL
IHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMA[RENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT
AKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQI
VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF
KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 7) (single underline: HNH domain; double underline: RuvC domain).

[00221] In some embodiments, the Cas9 domain comprises a D10A mutation, while the residue at position 840 remains a histidine in the amino acid sequence provided in SEQ ID NO: 6, or at corresponding positions in any of the amino acid sequences provided in another Cas9 domain, such as any of the Cas9 proteins provided herein. Without wishing to be bound by any particular theory, the presence of the catalytic residue H840 restores the acvitity of the Cas9 to cleave the non-edited *(e.g.,* non-deaminated) strand containing a G opposite the targeted C. Restoration of H840 *(e.g.,* from A840) does not result in the cleavage of the target strand containing the C. Such Cas9 variants are able to generate a single-strand DNA break (nick) at a specific location based on the gRNA-defined target sequence, leading to repair of the non-edited strand, ultimately resulting in a G to A change on the non-edited strand. A schematic representation of this process is shown in Figure 108. Briefly, the C of a C-G

basepair can be deaminated to a U by a deaminase, *e.g.*, an APOBEC deamonase. Nicking the non-edited strand, having the G, facilitates removal of the G via mismatch repair mechanisms. UGI inhibits UDG, which prevents removal of the U.

[00222] In other embodiments, dCas9 variants having mutations other than D10A and H840A are provided, which, *e.g.*, result in nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H820, or other substitutions within the nuclease domains of Cas9 *(e.g.*, substitutions in the HNH nuclease subdomain and/or the RuvCl subdomain). In some embodiments, variants or homologues of dCas9 *(e.g.*, variants of SEQ ID NO: 6) are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to SEQ ID NO: 6. In some embodiments, variants of dCas9 *(e.g.*, variants of SEQ ID NO: 6) are provided having amino acid sequences which are shorter, or longer than SEQ ID NO: 6, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids or more.

[00223] In some embodiments, Cas9 fusion proteins as provided herein comprise the full-length amino acid sequence of a Cas9 protein, *e.g.*, one of the Cas9 sequences provided herein. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length Cas9 sequence, but only a fragment thereof. For example, in some embodiments, a Cas9 fusion protein provided herein comprises a Cas9 fragment, wherein the fragment binds crRNA and tracrRNA or sgRNA, but does not comprise a functional nuclease domain, *e.g.*, in that it comprises only a truncated version of a nuclease domain or no nuclease domain at all. Exemplary amino acid sequences of suitable Cas9 domains and Cas9 fragments are provided herein, and additional suitable sequences of Cas9 domains and fragments will be apparent to those of skill in the art.

[00224] In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquis I* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref:

YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacterjejuni* (NCBI Ref: YP_002344900.1); or *Neisseria. meningitidis* (NCBI Ref: YP_002342100.1).

[00225] The term "deaminase" or "deaminase domain," as used herein, refers to a protein or enzyme that catalyzes a deamination reaction. In some embodiments, the deaminase or deaminase domain is a cytidine deaminase, catalyzing the hydrolytic deamination of cytidine or deoxycytidine to uridine or deoxyuridine, respectively. In some embodiments, the deaminase or deaminase domain is a cytidine deaminase domain, catalyzing the hydrolytic deamination of cytosine to uracil. In some embodiments, the deaminase or deaminase domain is a naturally-occuring deaminase from an organism, such as a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase or deaminase domain is a variant of a naturally-occuring deaminase from an organism, that does not occur in nature. For example, in some embodiments, the deaminase or deaminase domain is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75% at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occuring deaminase from an organism.

[00226] The term "effective amount," as used herein, refers to an amount of a biologically active agent that is sufficient to elicit a desired biological response. For example, in some embodiments, an effective amount of a nuclease may refer to the amount of the nuclease that is sufficient to induce cleavage of a target site specifically bound and cleaved by the nuclease. In some embodiments, an effective amount of a fusion protein provided herein, *e.g.,* of a fusion protein comprising a nuclease-inactive Cas9 domain and a nucleic acid editing domain *{e.g.,* a deaminase domain) may refer to the amount of the fusion protein that is sufficient to induce editing of a target site specifically bound and edited by the fusion protein. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.,* a fusion protein, a nuclease, a deaminase, a recombinase, a hybrid protein, a protein dimer, a complex of a protein (or protein dimer) and a polynucleotide, or a polynucleotide, may vary depending on various factors as, for example, on the desired biological response, *e.g.,* on the specific allele, genome, or target site to be edited, on the cell or tissue being targeted, and on the agent being used.

[00227] The term "linker," as used herein, refers to a chemical group or a molecule linking two molecules or moieties, *e.g.,* two domains of a fusion protein, such as, for example, a nuclease-inactive Cas9 domain and a nucleic acid editing domain *{e.g.,* a deaminase domain). A linker may be, for example, an amino acid sequence, a peptide, or a polymer of any length and compostion. In some embodiments, a linker joins a gRNA binding domain of an RNA-

programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of anucleic-acid editing protein. In some embodiments, a linker joins a dCas9 and a nucleic-acid editing protein. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids *(e.g.,* a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 1-100 amino acids in length, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated.

[00228] The term "mutation," as used herein, refers to a substitution of a residue within a sequence, *e.g.,* a nucleic acid or amino acid sequence, with another residue, or a deletion or insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Various methods for making the amino acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

[00229] The terms "nucleic acid" and "nucleic acid molecule," as used herein, refer to a compound comprising a nucleobase and an acidic moiety, *e.g.,* a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, *e.g.,* nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, "nucleic acid" refers to individual nucleic acid residues *(e.g.* nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably to refer to a polymer of nucleotides *(e.g.,* a string of at least three nucleotides). In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g., a* recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring

nucleotides or nucleosides. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, *e.g.,* analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, *e.g.,* in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5' to 3' direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides *(e.g.* adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs *(e.g.,* 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, 0(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases *(e.g.,* methylated bases); intercalated bases; modified sugars *(e.g.,* 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups *(e.g.,* phosphorothioates and 5′-*N*-phosphoramidite linkages).

[00230] The term "nucleic acid editing domain," as used herein refers to a protein or enzyme capable of making one or more modifications *(e.g.,* deamination of a cytidine residue) to a nucleic acid *(e.g.,* DNA or RNA). Exemplary nucleic acid editing domains include, but are not limited to a deaminase, a nuclease, a nickase, a recombinase, a methyltransferase, a methylase, an acetylase, an acetyltransferase, a transcriptional activator, or a transcriptional repressor domain. In some embodiments the nucleic acid editing domain is a deaminase *(e.g.,* a cytidine deaminase, such as an APOBEC or an AID deaminase).

[00231] The term "proliferative disease," as used herein, refers to any disease in which cell or tissue homeostasis is disturbed in that a cell or cell population exhibits an abnormally elevated proliferation rate. Proliferative diseases include hyperproliferative diseases, such as pre-neoplastic hyperplastic conditions and neoplastic diseases. Neoplastic diseases are characterized by an abnormal proliferation of cells and include both benign and malignant neoplasias. Malignant neoplasia is also referred to as cancer.

[00232] The terms "protein," "peptide," and "polypeptide" are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a

protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, *etc.* A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, recombinant, or synthetic, or any combination thereof. The term "fusion protein" as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins. One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an "amino-terminal fusion protein" or a "carboxy-terminal fusion protein," respectively. A protein may comprise different domains, for example, a nucleic acid binding domain *(e.g.,* the gRNA binding domain of Cas9 that directs the binding of the protein to a target site) and a nucleic acid cleavage domain or a catalytic domain of a nucleic-acid editing protein. In some embodiments, a protein comprises a proteinaceous part, *e.g.,* an amino acid sequence constituting a nucleic acid binding domain, and an organic compound, *e.g.,* a compound that can act as a nucleic acid cleavage agent. In some embodiments, a protein is in a complex with, or is in association with, a nucleic acid, *e.g.,* RNA. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

[00233] The term "subject," as used herein, refers to an individual organism, for example, an individual mammal. In some embodiments, the subject is a human. In some embodiments, the subject is a non-human mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the subject is a rodent. In some embodiments, the subject is a sheep, a goat, a cattle, a cat, or a dog. In some embodiments, the subject is a vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the subject is a research animal. In some embodiments, the subject is genetically engineered, *e.g.,* a

genetically engineered non-human subject. The subject may be of either sex and at any stage of development.

[00234] The term "target site" refers to a sequence within a nucleic acid molecule that is deaminated by a deaminase or a fusion protein comprising a deaminase, *(e.g.,* a dCas9-deaminase fusion protein provided herein).

[00235] The terms "treatment," "treat," and "treating" refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. As used herein, the terms "treatment," "treat," and "treating" refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In other embodiments, treatment may be administered in the absence of symptoms, *e.g.,* to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms *(e.g.,* in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their recurrence.

[00236] The term "recombinant" as used herein in the context of proteins or nucleic acids refers to proteins or nucleic acids that do not occur in nature, but are the product of human engineering. For example, in some embodiments, a recombinant protein or nucleic acid molecule comprises an amino acid or nucleotide sequence that comprises at least one, at least two, at least three, at least four, at least five, at least six, or at least seven mutations as compared to any naturally occurring sequence.

[00237] The term "pharmaceutical composition," as used herein, refers to a composition that can be administrated to a subject in the context of treatment of a disease or disorder. In some embodiments, a pharmaceutical composition comprises an active ingredient, e.g., a nuclease or a nucleic acid encoding a nuclease, and a pharmaceutically acceptable excipient.

[00238] The term "base editor (BE)," or "nucleobase editor (NBE)," as used herein, refers to an agent comprising a polypeptide that is capable of making a modification to a base *(e.g.,* A, T, C, G, or U) within a nucleic acid sequence *(e.g.,* DNA or RNA). In some embodiments, the base editor is capable of deaminating a base within a nucleic acid. In some embodiments, the base editor is capable of deaminating a base within a DNA molecule. In some embodiments, the base editor is capable of deaminating an cytosine (C) in DNA. In some embodiments, the

base editor is a fusion protein comprising a nucleic acid programmable DNA binding protein (napDNAbp) fused to a cytidine deaminase domain. In some embodiments, the base editor comprises a Cas9 *(e.g.,* dCas9 and nCas9), CasX, CasY, Cpfl, C2cl, C2c2, C2c3, or Argonaute protein fused to a cytidine deaminase. In some embodiments, the base editor comprises a Cas9 nickase (nCas9) fused to an cytidine deaminase. In some embodiments, the base editor comprises a nuclease-inactive Cas9 (dCas9) fused to a cytidine deaminase. In some embodiments, the base editor is fused to an inhibitor of base excision repair, for example, a UGI domain. In some embodiments, the base editor comprises a CasX protein fused to a cytidine deaminase. In some embodiments, the base editor comprises a CasY protein fused to a cytidine deaminase. In some embodiments, the base editor comprises a Cpfl protein fused to a cytidine deaminase. In some embodiments, the base editor comprises a C2cl protein fused to a cytidine deaminase. In some embodiments, the base editor comprises a C2c2 protein fused to a cytidine deaminase. In some embodiments, the base editor comprises a C2c3 protein fused to a cytidine deaminase. In some embodiments, the base editor comprises an Argonaute protein fused to a cytidine deaminase.

[00239] The term "uracil glycosylase inhibitor" or "UGI," as used herein, refers to a protein that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme.

[00240] The term "Cas9 nickase," as used herein, refers to a Cas9 protein that is capable of cleaving only one strand of a duplexed nucleic acid molecule *(e.g.,* a duplexed DNA molecule). In some embodiments, a Cas9 nickase comprises a D10A mutation and has a histidine at position H840 of SEQ ID NO: 6, or a corresponding mutation in another Cas9 domain, such as any of the Cas9 proteins provided herein. For example, a Cas9 nickase may comprise the amino acid sequence as set forth in SEQ ID NO: 8. Such a Cas9 nickase has an active HNH nuclease domain and is able to cleave the non-targeted strand of DNA, *i.e.,* the strand bound by the gRNA. Further, such a Cas9 nickase has an inactive RuvC nuclease domain and is not able to cleave the targeted strand of the DNA, *i.e.,* the strand where base editing is desired.

[00241] Exemplary Cas9 nickase (Cloning vector pPlatTET-gRNA2; Accession No. BAV54124).

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA

AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 8)

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION

[00242] Some aspects of this disclosure provide fusion proteins that comprise a domain capable of binding to a nucleotide sequence *(e.g.,* a Cas9, or a Cpfl protein) and an enzyme domain, for example, a DNA-editing domain, such as, *e.g.,* a deaminase domain. The deamination of a nucleobase by a deaminase can lead to a point mutation at the respective residue, which is referred to herein as nucleic acid editing. Fusion proteins comprising a Cas9 variant or domain and a DNA editing domain can thus be used for the targeted editing of nucleic acid sequences. Such fusion proteins are useful for targeted editing of DNA *in vitro, e.g.,* for the generation of mutant cells or animals; for the introduction of targeted mutations, *e.g.,* for the correction of genetic defects in cells *ex vivo, e.g.,* in cells obtained from a subject that are subsequently re-introduced into the same or another subject; and for the introduction of targeted mutations, *e.g.,* the correction of genetic defects or the introduction of deactivating mutations in disease-associated genes in a subject. Typically, the Cas9 domain of the fusion proteins described herein does not have any nuclease activity but instead is a Cas9 fragment or

a dCas9 protein or domain. Other aspects of the invention provide fusion proteins that comprise (i) a domain capable of binding to a nucleic acid sequence (e.g., a Cas9, or a Cpfl protein); (ii) an enzyme domain, for example, a DNA-editing domain (e.g., a deaminase domain); and (iii) one or more uracil glycosylase inhibitor (UGI) domains. The presence of at least one UGI domain increases base editing efficiency compared to fusion proteins without a UGI domain. A fusion protein comprising two UGI domains further increases base editing efficiency and product purity compared to fusion proteins with one UGI domain or without a UGI domain. Methods for the use of Cas9 fusion proteins as described herein are also provided.

Nucleic acid programmable DNA binding proteins

[00243] Some aspects of the disclosure provide nucleic acid programmable DNA binding proteins, which may be used to guide a protein, such as a base editor, to a specific nucleic acid *{e.g.,* DNA or RNA) sequence. It should be appreciated that any of the fusion proteins *{e.g.,* base editors) provided herein may include any nucleic acid programmable DNA binding protein (napDNAbp). For example, any of the fusion proteins described herein that include a Cas9 domain, can use another napDNAbp, such as CasX, CasY, Cpfl, C2cl, C2c2, C2c3, and Argonaute, in place of the Cas9 domain. Nucleic acid programmable DNA binding proteins include, without limitation, Cas9 *{e.g.,* dCas9 and nCas9), CasX, CasY, Cpfl, C2cl, C2c2, C2C3, and Argonaute. One example of a nucleic acid programmable DNA-binding protein that has a different PAM specificity than Cas9 is Clustered Regularly Interspaced Short Palindromic Repeats from Prevotella and Francisella 1 (Cpfl). Similar to Cas9, Cpfl is also a class 2 CRISPR effector. It has been shown that Cpflmediates robust DNA interference with features distinct from Cas9. Cpfl is a single RNA-guided endonuclease lacking tracrRNA, and it utilizes a T-rich protospacer-adjacent motif (TTN, TTTN, or YTN). Moreover, Cpfl cleaves DNA via a staggered DNA double-stranded break. Out of 16 Cpfl -family proteins, two enzymes from *Acidaminococcus* and *Lachnospiraceae* are shown to have efficient genome-editing activity in human cells. Cpfl proteins are known in the art and have been described previously, for example, Yamano *et al.,* "Crystal structure of Cpfl in complex with guide RNA and target DNA." *Cell* (165) 2016, p. 949-962; the entire contents of which are incorporated herein by reference.

[00244] Also useful in the present compositions and methods are nuclease-inactive Cpfl (dCpfl) variants that may be used as a guide nucleotide sequence-programmable DNA-binding protein domain. The Cpfl protein has a RuvC-like endonuclease domain that is

similar to the RuvC domain of Cas9 but does not have a HNH endonuclease domain, and the N-terminal of Cpfl does not have the alpha-helical recognition lobe of Cas9. It was shown in Zetsche *et al., Cell,* 163, 759-771, 2015 (which is incorporated herein by reference) that, the RuvC-like domain of Cpfl is responsible for cleaving both DNA strands and inactivation of the RuvC-like domain inactivates Cpfl nuclease activity. For example, mutations corresponding to D917A, E1006A, or D1255A in *Francisella novicida* Cpfl (SEQ ID NO: 15) inactivate Cpfl nuclease activity. In some embodiments, the dead Cpfl (dCpfl) comprises mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or D917A/E1006A/D1255A in SEQ ID NO: 9. It is to be understood that any mutations, *e.g.,* substitution mutations, deletions, or insertions, that inactivate the RuvC domain of Cpfl, may be used in accordance with the present disclosure.

[00245] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein is a Cpfl protein. In some embodiments, the Cpfl protein is a Cpfl nickase (nCpfl). In some embodiments, the Cpfl protein is a nuclease inactive Cpfl (dCpfl). In some embodiments, the Cpfl, the nCpfl, or the dCpfl comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of SEQ ID NOs: 9-24. In some embodiments, the dCpfl comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of SEQ ID NOs: 9-16, and comprises mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or D917A/ E1006A/D1255A in SEQ ID NO: 9. In some embodiments, the dCpf 1 protein comprises an amino acid sequence of any one SEQ ID NOs: 9-16. It should be appreciated that Cpfl from other species may also be used in accordance with the present disclosure.

Wild type *Francisella novicida* Cpfl (SEQ ID NO: 9) (D917, E1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKD YKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK

RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 9)

Francisella novicida Cpfl D917A (SEQ ID NO: 10) (A917, E1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF

ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 10)

Francisella novicida Cpfl E1006A (SEQ ID NO: 11) (D917, A1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI

SPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 11)

Francisella novicida Cpfl D1255A (SEQ ID NO: 12) (D917, E1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKD YKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 12)

Francisella novicida Cpfl D917A/E1006A (SEQ ID NO: 13) (A917, A1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKD YKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY

FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 13)

Francisella novicida Cpf1 D917A/D1255A (SEQ ID NO: 14) (A917, E1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN

KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 14)

Francisella novicida Cpf1 E1006A/D1255A (SEQ ID NO: 15) (D917, A1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK

ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 15)

Francisella novicida Cpfl D917A/E1006A/D1255A (SEQ ID NO: 16) (A917, A1006, and A1255 are bolded and underlined)
MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKD YKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 16)

[00246] In some embodiments, the nucleic acid programmable DNA binding protein is a Cpfl protein from an *Acidaminococcus* species (AsCpfl). Cpfl proteins form *Acidaminococcus* species have been described previously and would be apparent to the skilled

artisan. Exemplary *Acidaminococcus* Cpfl proteins (AsCpfl) include, without limitation, any of the AsCpfl proteins provided herin

[00247] Wild-type AsCpfl- Residue R912 is indicated in bold underlining and residues 661-667 are indicated in italics and underlining.

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD
AINKRHAEIYKGLFKAELFNGKVLKQLGT VTTTEHENALLRSFDKFTTYFSGFYENRK
NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEE
VFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLP
HRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLY
HLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNF
QMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFD
KMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEK
EPKKFQTAYA^TG^fi^GYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLG
EYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWT
GLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLY
QELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAA
NSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKK
LDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFK
SKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGT
QSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILH
FKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYR
DLYPANELIALLEEKGIVFRDGS NILPKLLENDDS HAIDTMVALIRS VLQMRNS NAATG
EDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKL
QNGIS NQDWLAYIQELRN (SEQ ID NO: 17)

[00248] AsCpfl (R912A)- Residue A912 is indicated in bold underlining and residues 661-667 are indicated in italics and underlining.

[00249] TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPII
DRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIG
RTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFS

GFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGI
FVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDET
AHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALF
NELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSL
KHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQL
DSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVE
KFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALS FEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYD
LNNPEKEPKKFQTAYA^TG^fi^GYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSS
QYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNL
HTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKT
PIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPIT
LNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGEANLIYITVIDSTGKILEQRSLNTIQQ
FDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLE
NLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFT
SFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDV
KTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIEN
HRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQM
RNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHL
KESKDLKLQNGISNQDWLAYIQELRN (SEQ ID NO: 19)

[00250] In some embodiments, the nucleic acid programmable DNA binding protein is a Cpf1 protein from a *Lachnospiraceae* species (LbCpf1). Cpf1 proteins form *Lachnospiraceae* species have been described previously have been described previously and would be apparent to the skilled artisan. Exemplary *Lachnospiraceae* Cpfl proteins (LbCpfl) include, without limitation, any of the LbCpfl proteins provided herein.

[00251] In some embodiments, the LbCpfl is a nickase. In some embodiments, the LbCpfl nickase comprises an R836X mutant relative to SEQ ID NO: 18, wherein X is any amino acid except for R. In some embodiments, the LbCpfl nickase comprises R836A mutant relative to SEQ ID NO: 18. In some embodiments, the LbCpfl is a nuclease inactive LbCpfl (dLbCpfl). In some embodiments, the dLbCpfl comprises a D832X mutant relative to SEQ ID NO: 18, wherein X is any amino acid except for D. In some embodiments, the dLbCpfl comprises a D832A mutant relative to SEQ ID NO: 18. Additional dCpfl proteins have been described in the art, for example, in Li *et al.* "Base editing with a Cpfl-cytidine deaminase fusion" *Nature*

Biotechnology; March 2018 DOI: 10.1038/nbt.4102; the entire contents of which are incorporated herein by reference. In some embodiments, the dCpfl comprises 1, 2, or 3 of the point mutations D832A, E1006A, Dl 125A of the Cpfl described in *Li et al.*

[00252] Wild-type LbCpf1 - Residues R836 and Rl 138 is indicated in bold underlining.
MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRY
YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYK
SLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCIN
ENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDV
YNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYT
SDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEK
LKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKA
FFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQF
MGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKL
LPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDS
ISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKL
YMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEEL
VVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTE
VRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHS
LLDKKEKERFEARQNWTSffiNIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFK
NSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMST
QNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALD
YKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGIN
YQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYD
SRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLE
YAQTSVKH (SEQ ID NO: 18)

[00253] LbCpfl (R836A)- Residue A836 is indicated in bold underlining.
MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRY
YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYK
SLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCIN
ENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDV
YNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYT

SDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEK
LKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKA
FFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQF
MGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKL
LPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDS
ISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKL
YMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEEL
VVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTE
VRVLLKHDDNPYVIGIDRGEANLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHS
LLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFK
NSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMST
QNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALD
YKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGIN
YQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYD
SRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLE
YAQTSVKH (SEQ ID NO: 20)

[00254] LbCpf1 (R1138A)- Residue A1138 is indicated in bold underlining.

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRY
YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYK
SLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCIN
ENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDV
YNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYT
SDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEK
LKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKA
FFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQF
MGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKL
LPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDS
ISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKL
YMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEEL
VVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTE
VRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHS

LLDKKEKERFEARQNWTSffiNIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFK NSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMST QNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALD YKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGIN YQQGDIRALLCEQSDKAFYSSFMALMSLMLQMANSITGRTDVDFLISPVKNSDGIFYD SRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLE YAQTSVKH (SEQ ID NO: 21)

[00255] In some embodiments, the Cpfl protein is a crippled Cpf 1 protein. As used herein a "crippled Cpfl" protein is a Cpfl protein having diminished nuclease activity as compared to a wild-type Cpfl protein. In some embodiments, the crippled Cpfl protein preferentially cuts the target strand more efficiently than the non-target strand. For example, the Cpfl protein preferentially cuts the strand of a duplexed nucleic acid molecule in which a nucleotide to be edited resides. In some embodiments, the crippled Cpfl protein preferentially cuts the non-target strand more efficiently than the target strand. For example, the Cpfl protein preferentially cuts the strand of a duplexed nucleic acid molecule in which a nucleotide to be edited does not reside. In some embodiments, the crippled Cpfl protein preferentially cuts the target strand at least 5% more efficiently than it cuts the non-target strand. In some embodiments, the crippled Cpfl protein preferentially cuts the target strand at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, 60%, 70%, 80%, 90%, or at least 100% more efficiently than it cuts the non-target strand.

[00256] In some embodiments, a crippled Cpfl protein is a non-naturally occurring Cpfl protein. In some embodiments, the crippled Cpfl protein comprises one or more mutations relative to a wild-type Cpfl protein. In some embodiments, the crippled Cpfl protein comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 mutations relative to a wild-type Cpfl protein. In some embodiments, the crippled Cpfl protein comprises an R836A mutation mutation as set forth in SEQ ID NO: 18, or in a corresponding amino acid in another Cpfl protein. It should be appreciated that a Cpfl comprising a homologous residue *(e.g.,* a corresponding amino acid) to R836A of SEQ ID NO: 18 could also be mutated to achieve similar results. In some embodiments, the crippled Cpfl protein comprises a R1138A mutation as set forth in SEQ ID NO: 18, or in a corresponding amino acid in another Cpfl protein. In some embodiments, the crippled Cpfl protein comprises an R912A mutation mutation as set forth in SEQ ID NO: 17, or in a corresponding amino acid in another Cpfl protein. Without wishing to be bound by any particular theory, residue R836 of SEQ ID NO:

18 (LbCpf1) and residue R912 of SEQ ID NO: 17 (AsCpf1) are examples of corresponding *(e.g.,* homologous) residues. For example, a portion of the alignment between SEQ ID NO: 17 and 18 shows that R912 and R836 are corresponding residues.

```
AsCpf1      YQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQ--
LbCpf1      KCPKN-IFKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQVSLNEIINN
                 *   *:* .*.. **.. :  :**********:**.*:*.:*:*:** *** *
```

[00257] In some embodiments, any of the Cpf1 proteins provided herein comprises one or more amino acid deletions. In some embodiments, any of the Cpf1 proteins provided herein comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acid deletions. Without wishing to be bound by any particular theory, there is a helical region in Cpf1, which includes residues 661-667 of AsCpf1 (SEQ ID NO: 17), that may obstruct the function of a deaminase *(e.g.,* APOBEC) that is fused to the Cpf1. This region comprises the amino acid sequence KKTGDQK. Accordingly, aspects of the disclosure provide Cpf1 proteins comprising mutations *(e.g.,* deletions) that disrupt this helical region in Cpf1. In some embodiments, the Cpf1 protein comprises one or more deletions of the following residues in SEQ ID NO: 17, or one or more corresponding deletions in another Cpf1 protein: K661, K662, T663, G664, D665, Q666, and K667. In some embodiments, the Cpf1 protein comprises a T663 and a D665 deletion in SEQ ID NO: 17, or corresponding deletions in another Cpf1 protein. In some embodiments, the Cpf1 protein comprises a K662,T663, D665, and Q666 deletion in SEQ ID NO: 17, or corresponding deletions in another Cpf1 protein. In some embodiments, the Cpf1 protein comprises a K661, K662, T663, D665, Q666 and K667 deletion in SEQ ID NO: 17, or corresponding deletions in another Cpf1 protein.

[00258] AsCpf1 (deleted T663 and D665)

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD
AINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRK
NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEE
VFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLP
HRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLY
HLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNF
QMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFD

KMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEK
EPKKFQTAYAKKGQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEY
YAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGL
FSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQE
LYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANS
PSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLD
NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSK
RTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQS
GFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFK
MNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDL
YPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGED
YINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQN
GISNQDWLAYIQELRN (SEQ ID NO: 22)

[00259] AsCpfl (deleted K662, T663, D665, and Q666)
TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD
AINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRK
NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEE
VFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLP
HRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLY
HLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNF
QMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFD
KMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEK
EPKKFQTAYAKGKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYA
ELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFS
PENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELY
DYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPS
KFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKR
TGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQS
GFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFK

MNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDL
YPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGED
YINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQN
GISNQDWLAYIQELRN (SEQ ID NO: 23)

[00260] AsCpfl (deleted K661, K662, T663,D665, Q666, and K667)
TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD
AINKRHAEIYKGLFKAELFNGKVLKQLGT VTTTEHENALLRSFDKFTTYFSGFYENRK
NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEE
VFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLP
HRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLY
HLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNF
QMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFD
KMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEK
EPKKFQTAYAGGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAEL
NPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPE
NLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYD
YVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKF
NQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNRE
KERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTG
IAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFL
FYVPAPYTSKIDPLTGFVDPFVWKTIKNHES RKHFLEGFDFLHYD VKTGDFILHFKMN
RNLSFQRGLPGFMPAWDrVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYP
ANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYI
NSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGIS
NQDWLAYIQELRN (SEQ ID NO: 24)

[00261] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a nucleic acid programmable DNA binding protein that does not require a canonical (NGG) PAM sequence in the target sequence. In some embodiments, the napDNAbp is an Argonaute protein. One example of such a nucleic acid programmable DNA

binding protein is an Argonaute protein from *Natronobacterium gregoryi* (NgAgo). NgAgo is a ssDNA-guided endonuclease. NgAgo binds 5'-phosphorylated ssDNA of ~24 nucleotides (gDNA) in length to guide it to a target site and makes DNA double-strand breaks at the gDNA site. In contrast to Cas9, the NgAgo-gDNA system does not require a protospacer-adjacent motif (PAM). Using a nuclease inactive NgAgo (dNgAgo) can greatly expand the bases that may be targeted. The characterization and use of NgAgo have been described in Gao *et al., Nat. Biotechnol.,* 2016 Jul;34(7):768-73. PubMed PMID: 27136078; Swarts *et al., Nature* 507(7491) (2014):258-61; and Swarts *et al, Nucleic Acids Res.* 43(10) (2015):5120-9, each of which is incorporated herein by reference. The sequence of *Natronobacterium gregoryi* Argonaute is provided in SEQ ID NO: 25.

[00262] In some embodiments, the napDNAbp is an Argonaute protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Argonaute protein. In some embodiments, the napDNAbp is a naturally-occurring Argonaute protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of SEQ ID NO: 25. In some embodiments, the napDNAbp comprises an amino acid sequence of any one SEQ ID NO: 25.

Wild type *Natronobacterium gregoryi* Argonaute (SEQ ID NO: 25)
MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYITLW
KNTTPKDVFTYDYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTDED
ETFAGGEPLDHHLDDALNETPDDAETESDSGHVMTSFASRDQLPEWTLHTYTLTATD
GAKTDTEYARRTLAYTVRQELYTDHDAAPVATDGLMLLTPEPLGETPLDLDCGVRV
EADETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLSKEPVLTCDEFDL
HERYDLSVEVGHSGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRVKTTYRPRRGHIV
WGLRDECATDSLNTLGNQSVVAYHRNNQTPINTDLLDAIEAADRRVVETRRQGHGD
DAVSFPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEKAQAFAERLDPVRL
NGSTVEFSSEFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDETFSKGIVNPPESFEV
AVVLPEQQADTCKAQWDTMADLLNQAGAPPTRSETVQYDAFSSPESISLNVAGAIDP
SEVDAAFVVLPPDQEGFADLASPTETYDELKKALANMGIYSQMAYFDRFRDAKIFYT
RNVALGLLAAAGGVAFTTEHAMPGDADMFIGIDVSRSYPEDGASGQINIAATATAVY

KDGTILGHSSTRPQLGEKLQSTDVRDIMKNAILGYQQVTGESPTHIVIHRDGFMNEDL DPATEFLNEQGVEYDIVEIRKQPQTRLLAVSDVQYDTPVKSIAAINQNEPRATVATFG APEYLATRDGGGLPRPIQIERVAGETDIETLTRQVYLLSQSHIQVHNSTARLPITTAYA DQASTHATKGYLVQTGAFESNVGFL (SEQ ID NO: 25)

[00263] In some embodiments, the napDNAbp is a prokaryotic homolog of an Argonaute protein. Prokaryotic homologs of Argonaute proteins are known and have been described, for example, in Makarova K., *et al.,* "Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements", *Biol. Direct.* 2009 Aug 25;4:29. doi: 10.1186/1745-6150-4-29, is incorporated herein by reference. In some embodiments, the napDNAbp is a *Marinitoga piezophila* Argunaute (MpAgo) protein. The CRISPR-associated *Marinitoga piezophila* Argonaute (MpAgo) protein cleaves single-stranded target sequences using 5'-phosphorylated guides. The 5' guides are used by all known Argonautes. The crystal structure of an MpAgo-RNA complex shows a guide strand binding site comprising residues that block 5' phosphate interactions. This data suggests the evolution of an Argonaute subclass with noncanonical specificity for a 5'-hydroxylated guide. See, *e.g.,* Kaya *et al.,* "A bacterial Argonaute with noncanonical guide RNA specificity", *Proc Natl Acad Sci USA*. 2016 Apr 12;113(15):4057-62, the entire contents of which are hereby incorporated by reference). It should be appreciated that other Argonaute proteins may be used in any of the fusion proteins *{e.g.,* base editors) described herein, for example, to guide a deaminase *{e.g.,* cytidine deaminase) to a target nucleic acid *{e.g.,* ssRNA).

[00264] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a single effector of a microbial CRISPR-Cas system. Single effectors of microbial CRISPR-Cas systems include, without limitation, Cas9, Cpfl, C2cl, C2c2, and C2c3. Typically, microbial CRISPR-Cas systems are divided into Class 1 and Class 2 systems. Class 1 systems have multisubunit effector complexes, while Class 2 systems have a single protein effector. Cas9 and Cpfl are Class 2 effectors. In addition to Cas9 and Cpfl, three distinct Class 2 CRISPR-Cas systems (C2cl, C2c2, and C2c3) have been described by Shmakov *et al.,* "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", *Mol. Cell,* 2015 Nov 5; 60(3): 385-397, the entire contents of which are herein incorporated by reference. Effectors of two of the systems, C2cl and C2c3, contain RuvC-like endonuclease domains related to Cpfl. A third system, C2c2 contains an effector with two predicted HEPN RNase domains. Production of mature CRISPR RNA is tracrRNA-

independent, unlike production of CRISPR RNA by C2cl. C2cl depends on both CRISPR RNA and tracrRNA for DNA cleavage. Bacterial C2c2 has been shown to possess a unique RNase activity for CRISPR RNA maturation distinct from its RNA-activated single-stranded RNA degradation activity. These RNase functions are different from each other and from the CRISPR RNA-processing behavior of Cpfl. See, *e.g.,* East-Seletsky, *et al.,* "Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection", *Nature,* 2016 Oct 13;538(7624):270-273, the entire contents of which are hereby incorporated by reference. *In vitro* biochemical analysis of C2c2 in *Leptotrichia shahii* has shown that C2c2 is guided by a single CRISPR RNA and can be programmed to cleave ssRNA targets carrying complementary protospacers. Catalytic residues in the two conserved HEPN domains mediate cleavage. Mutations in the catalytic residues generate catalytically inactive RNA-binding proteins. See *e.g.,* Abudayyeh *et al.,* "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector," *Science,* 2016 Aug 5; 353(6299), the entire contents of which are hereby incorporated by reference.

[00265] The crystal structure of *Alicyclobaccillus acidoterrastris* C2cl (AacC2cl) has been reported in complex with a chimeric single-molecule guide RNA (sgRNA). See, *e.g.,* Liu *et al.,* "C2cl-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", *Mol. Cell,* 2017 Jan 19;65(2):310-322, incorporated herein by reference. The crystal structure has also been reported for *Alicyclobacillus acidoterrestris* C2cl bound to target DNAs as ternary complexes. See, *e.g.,* Yang *et al.,* "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell,* 2016 Dec 15;167(7):1814-1828, the entire contents of which are hereby incorporated by reference. Catalytically competent conformations of AacC2cl, both with target and non-target DNA strands, have been captured independently positioned within a single RuvC catalytic pocket, with C2cl -mediated cleavage resulting in a staggered seven-nucleotide break of target DNA. Structural comparisons between C2cl ternary complexes and previously identified Cas9 and Cpfl counterparts demonstrate the diversity of mechanisms used by CRISPR-Cas9 systems.

[00266] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein is a C2cl, a C2c2, or a C2c3 protein. In some embodiments, the napDNAbp is a C2cl protein. In some embodiments, the napDNAbp is a C2c2 protein. In some embodiments, the napDNAbp is a C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring C2cl,

C2c2, or C2c3 protein. In some embodiments, the napDNAbp is a naturally-occurring C2cl, C2c2, or C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of SEQ ID NOs: 26-28. In some embodiments, the napDNAbp comprises an amino acid sequence of any one SEQ ID NOs: 26-28. It should be appreciated that C2cl, C2c2, or C2c3 from other bacterial species may also be used in accordance with the present disclosure.

C2c 1 (uniprot.org/uniprot/T0D7 A2#)

splT0D7A2IC2Cl_ALIAG CRISPR-associated endonuclease C2cl *OS=Alicyclobacillus acidoterrestris* (strain ATCC 49025 / DSM 3922 / CIP 106132 / NCIMB 13137 / GD3B) GN=c2cl PE=1 SV=1

MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGD
GEQECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGA
KGDAQQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETR
KSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQA
IERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASP
GLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGS
HDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPrWT
RFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDN
LLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSV
RVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSG
LRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLLK
LPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVD
AANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDW
RKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGS
RFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGKWVAKYPPCQLILLEE
LSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDART
GAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPF
SAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTA
DSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAELLVEADEAREKSVVLM
RDPSGIINRGNWTRQKEFWSMVNQRIEGYLVKQIRSRVPLQDSACENTGDI (SEQ ID NO: 26)

C2c2 (uniprot.org/uniprot/P0DOC6)
>splP0DOC6IC2C2_LEPSD CRISPR-associated endoribonuclease C2c2 OS=Leptotrichia shahii (strain DSM 19757 / CCUG 47503 / CIP 107916 / JCM 16776 / LB37) GN=c2c2 PE=1 SV=1

[00267] MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNNKEKID NNKFIRKYINYKKNDNILKEFTRKFHAGNILFKLKGKEGIIRIENNDDFLETEEVVLYIE AYGKSEKLKALGITKKKIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIRDEYTNKTLNDC SIILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRYYEEHLREKLLKDD KIDVILTNFMEIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVEKILNINVDLTVE DIADFVIKELEFWNITKRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKK DKIVKFFVENIKNNSIKEKIEKILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHYKVNF DSKKFSKKSDEEKELYKIIYRYLKGRIEKILVNEQKVRLKKMEKIEIEKILNESILSEKIL KRVKQYTLEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLELITFFASTNMELN KIFSRENINNDENIDFFGGDREKNYVLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFTK IGTNERNRILHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKKNII TKINDIKISEENNNDIKYLPSFSKVLPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKE LYKKLILEDDLEENES KNIFLQELKKTLGNIDEIDENIIENYYKNAQIS ASKGNNKAIKK YQKKVIECYIGYLRKNYEELFDFSDFKMNIQEIKKQIKDINDNKTYERITVKTSDKTIVI NDDFEYIISIFALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTLRNECIT ENWNLNLEEFIQKMKEIEKDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVLE KKLEKIVIFDDETKFEIDKKSNILQDEQRKLSNINKKDLKKKVDQYIKDKDQEIKSKIL CRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIYYPKERKNELYIYKKNLFLNIGNPN FDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNGYSKEYKEKYI KKLKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIESYLIDINWKL AIQMARFERDMHYIVNGLRELGIIKLSGYNTGISRAYPKRNGSDGFYTTTAYYKFFDE ESYKKFEKIC YGFGIDLS ENS EINKPENES IRNYIS HFYIVRNPF ADY SIAEQIDRVS NLLS YSTRYNNS TYA SVFE VFKKD VNLD YDELKKKFKLIGNNDILERLMKPKKVS VLELES YNSDYIKNLIIELLTKIENTNDTL (SEQ ID NO: 27)

[00268] C2c3, translated from >CEPX01008730.1 marine metagenome genome assembly TARA_037_MES_0. 1-0.22, contig TARA_037_MES_0.1-0.22_scaffold22115_l, whole genome shotgun sequence.

[00269] MRSNYHGGRNARQWRKQISGLARRTKETVFTYKFPLETDAAEIDFDKAVQ TYGIAEGVGHGSLIGLVCAFHLSGFRLFSKAGEAMAFRNRSRYPTDAFAEKLSAIMGI QLPTLSPEGLDLIFQSPPRSRDGIAPVWSENEVRNRLYTNWTGRGPANKPDEHLLEIAG EIAKQVFPKFGGWDDLASDPDKALAAADKYFQSQGDFPSIASLPAAHVILSPANSTVDF EGDYIAIDPAAETLLHQAVSRCAARLGRERPDLDQNKGPFVSSLQDALVSSQNNGLS WLFGVGFQHWKEKSPKELIDEYKVPADQHGAVTQVKSFVDAIPLNPLFDTTHYGEFR ASVAGKVRSWVANYWKRLLDLKSLLATTEFTLPESISDPKAVSLFSGLLVDPQGLKK VADSLPARLVSAEEAIDRLMGVGIPTAADIAQVERVADEIGAFIGQVQQFNNQVKQKL ENLQDADDEEFLKGLKIELPSGDKEPPAINRISGGAPDAAAEISELEEKLQRLLDARSE HFQTISEWAEENAVTLDPIAAMVELERLRLAERGATGDPEEYALRLLLQRIGRLANRV SPVSAGSIRELLKPVFMEEREFNLFFHNRLGS LYRSPYSTSRHQPFSIDVGKAKAIDWIA GLDQIS SDIEKALS GAGEALGDQLRDWINLAGFAIS QRLRGLPDTVPNALAQVRCPDD VRIPPLLAMLLEEDDIARDVCLKAFNLYVSAINGCLFGALREGFIVRTRFQRIGTDQIH YVPKDKAWEYPDRLNTAKGPINAAVSSDWIEKDGAVIKPVETVRNLSSTGFAGAGVS EYLVQAPHDWYTPLDLRDVAHLVTGLPVEKNITKLKRLTNRTAFRMVGASSFKTHLD SVLLSDKIKLGDFTIIIDQHYRQSVTYGGKVKISYEPERLQVEAAVPVVDTRDRTVPEP DTLFDHIVAIDLGERSVGFAVFDIKSCLRTGEVKPIHDNNGNPVVGTVAVPSIRRLMK AVRSHRRRRQPNQKVNQTYSTALQNYRENVIGDVCNRIDTLMERYNAFPVLEFQIKN FQAGAKQLEIVYGS (SEQ ID NO: 28)

[00270] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein is a Cas9 from archaea *(e.g.* nanoarchaea), which constitute a domain and kingdom of single-celled prokaryotic microbes. In some embodiments, the napDNAbp is CasX or CasY, which have been described in, for example, Burstein *et al.,* "New CRISPR-Cas systems from uncultivated microbes." *Cell Res.* 2017 Feb 21. doi: 10.1038/cr.2017.21, which is incorporated herein by reference. Using genome-resolved metagenomics, a number of CRISPR-Cas systems were identified, including the first reported Cas9 in the archaeal domain of life. This divergent Cas9 protein was found in nanoarchaea as part of an active CRISPR-Cas system. In bacteria, two previously unknown systems were discovered, CRISPR-CasX and CRISPR-CasY, which are among the most compact systems yet discovered. In some embodiments, Cas9 refers to CasX, or a variant of CasX. In some embodiments, Cas9 refers to a CasY, or a variant of CasY. It should be appreciated that other RNA-guided DNA binding proteins may be used as a nucleic acid programmable DNA binding protein (napDNAbp) and are within the scope of this disclosure.

[00271] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein is a CasX or CasY protein. In some embodiments, the napDNAbp is a CasX protein. In some embodiments, the napDNAbp is a CasY protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring CasX or CasY protein. In some embodiments, the napDNAbp is a naturally-occurring CasX or CasY protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of SEQ ID NOs: 29-31. In some embodiments, the napDNAbp comprises an amino acid sequence of any one SEQ ID NOs: 29-31. It should be appreciated that CasX and CasY from other bacterial species may also be used in accordance with the present disclosure.

CasX (uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53)
>trlF0NN87IF0NN87_SULIH CRISPR-associated Casx protein *OS=Sulfolobus islandicus* (strain HVE10/4) GN=SiH_0402 PE=4 SV=1
MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAE
RRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVK
ECEEVSAPSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVS
EGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRI
YTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANY
IYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG (SEQ ID NO: 29)

>trlF0NH53IF0NH53_SULIR CRISPR associated protein, Casx *OS=Sulfolobus islandicus* (strain REY15A) GN=SiRe_0771 PE=4 SV=1
MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAE
RRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVK
ECEEVSAPSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKV
SEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVS
IYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLAN

YIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG (SEQ ID NO: 30)

CasY (ncbi.nlm.nih.gov/protein/APG80656. 1)
>APG80656.1 CRISPR-associated protein CasY [uncultured Parcubacteria group bacterium]

[00272] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIV SAINDDYVGLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLK NVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFK AEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNP LNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFS NFLGEGFLGRLRENKITELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKF DNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTK EEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIK ERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKK YKNAAIYTDALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDK **WKPrVKNS**FAPYCDIVSLAENEVLYKPKQS**RSRKS**AAIDKNR**VRLPS**TENIAKAGIALA RELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVENGTVK DFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELL YIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYE LTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQF LEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPFTI FPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKL DQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKV YATLKKADVYSEIDADKNLQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVD ETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHISKKMRGNS CLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI (SEQ ID NO: 31)

Cas9 domains of Nucleobase Editors

[00273] Non-limiting, exemplary Cas9 domains are provided herein. The Cas9 domain may be a nuclease active Cas9 domain, a nucleasae inactive Cas9 domain, or a Cas9 nickase. In some embodiments, the Cas9 domain is a nuclease active domain. For example, the Cas9 domain may be a Cas9 domain that cuts both strands of a duplexed nucleic acid *{e.g.,* both strands of a duplexed DNA molecule). In some embodiments, the Cas9 domain comprises any

Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments the Cas9 domain comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the Cas9 domain comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more mutations compared to any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the Cas9 domain comprises an amino acid sequence that has at least 10, at least 15, at least 20, at least 30, at leat 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, or at least 1200 identical contiguous amino acid residues as compared to any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein.

[00274] In some embodiments, the Cas9 domain is a nuclease-inactive Cas9 domain (dCas9). For example, the dCas9 domain may bind to a duplexed nucleic acid molecule *(e.g.,* via a gRNA molecule) without cleaving either strand of the duplexed nucleic acid molecule. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10X mutation and a H840X mutation of the amino acid sequence set forth in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, wherein X is any amino acid change. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10A mutation and a H840A mutation of the amino acid sequence set forth in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. As one example, a nuclease-inactive Cas9 domain comprises the amino acid sequence set forth in SEQ ID NO: 32 (Cloning vector pPlatTET-gRNA2, Accession No. BAV54124).

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGAS QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS

IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEE VVDKGASAQSFIERMTNFD KNLPNEKVLPKHS LLYEYFTVYNELTK V
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKS EQEIGKAT AKYFFYS NIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 32; see, *e.g.*, Qi *et al*, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* 2013; 152(5): 1173-83, the entire contents of which are incorporated herein by reference).

[00275] Additional suitable nuclease-inactive dCas9 domains will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A/H840A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (See, *e.g.*, Prashant *et al*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838, the entire contents of which are incorporated herein by reference). In some embodiments the dCas9 domain comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the dCas9 domains provided herein. In some embodiments, the Cas9 domain comprises an amino acid sequences that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more or more mutations

compared to any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the Cas9 domain comprises an amino acid sequence that has at least 10, at least 15, at least 20, at least 30, at leat 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, or at least 1200 identical contiguous amino acid residues as compared to any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein.

[00276] In some embodiments, the Cas9 domain is a Cas9 nickase. The Cas9 nickase may be a Cas9 protein that is capable of cleaving only one strand of a duplexed nucleic acid molecule *(e.g.,* a duplexed DNA molecule). In some embodiments the Cas9 nickase cleaves the target strand of a duplexed nucleic acid molecule, meaning that the Cas9 nickase cleaves the strand that is base paired to (complementary to) a gRNA *(e.g.,* an sgRNA) that is bound to the Cas9. In some embodiments, a Cas9 nickase comprises a D10A mutation and has a histidine at position 840 of SEQ ID NO: 6, or a mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. For example, a Cas9 nickase may comprise the amino acid sequence as set forth in SEQ ID NO: 8. In some embodiments the Cas9 nickase cleaves the non-target, non-base-edited strand of a duplexed nucleic acid molecule, meaning that the Cas9 nickase cleaves the strand that is not base paired to a gRNA *(e.g.,* an sgRNA) that is bound to the Cas9. In some embodiments, a Cas9 nickase comprises an H840A mutation and has an aspartic acid residue at position 10 of SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments the Cas9 nickase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the Cas9 nickases provided herein. Additional suitable Cas9 nickases will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure.

Cas9 Domains with Reduced PAM Exclusivity

[00277] Some aspects of the disclosure provide Cas9 domains that have different PAM specificities. Typically, Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a canonical NGG PAM sequence to bind a particular nucleic acid region. This may limit the ability to edit desired bases within a genome. In some embodiments, the base editing fusion proteins provided herein may need to be placed at a precise location, for example where a

target base is placed within a 4 base region *(e.g.,* a "deamination window"), which is approximately 15 bases upstream of the PAM. See Komor, A.C., *et al.,* "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016), the entire contents of which are hereby incorporated by reference. Accordingly, in some embodiments, any of the fusion proteins provided herein may contain a Cas9 domain that is capable of binding a nucleotide sequence that does not contain a canonical *(e.g.,* NGG) PAM sequence. Cas9 domains that bind to non-canonical PAM sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.,* "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.,* "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology* 33, 1293-1298 (2015); the entire contents of each are hereby incorporated by reference.

[00278] In some embodiments, the Cas9 domain is a Cas9 domain from *Staphylococcus aureus* (SaCas9). In some embodiments, the SaCas9 domain is a nuclease active SaCas9, a nuclease inactive SaCas9 (SaCas9d), or a SaCas9 nickase (SaCas9n). In some embodiments, the SaCas9 comprises the amino acid sequence SEQ ID NO: 33. In some embodiments, the SaCas9 comprises a N579X mutation of SEQ ID NO: 33, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, wherein X is any amino acid except for N. In some embodiments, the SaCas9 comprises a N579A mutation of SEQ ID NO: 33, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid seuqnce having a non-canonical PAM. In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a NNGRRT PAM sequence. In some embodiments, the SaCas9 domain comprises one or more of a E781X, a N967X, and a R1014X mutation of SEQ ID NO: 33, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, wherein X is any amino acid. In some embodiments, the SaCas9 domain comprises one or more of a E781K, a N967K, and a R1014H mutation of SEQ ID NO: 33, or one or more corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the SaCas9 domain comprises a E781K, a N967K, or a R1014H mutation of SEQ ID NO: 33, or corresponding mutations in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein.

[00279] In some embodiments, the Cas9 domain of any of the fusion proteins provided herein comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of SEQ ID NOs: 33-36. In some embodiments, the Cas9 domain of any of the fusion proteins provided herein comprises the amino acid sequence of any one of SEQ ID NOs: 33-36. In some embodiments, the Cas9 domain of any of the fusion proteins provided herein consists of the amino acid sequence of any one of SEQ ID NOs: 33-36.

Exemplary SaCas9 sequence

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERffiEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD
TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G (SEQ ID NO: 33)

Residue N579 of SEQ ID NO: 33, which is underlined and in bold, may be mutated *(e.g.,* to a A579) to yield a SaCas9 nickase.

Exemplary SaCas9d sequence

KRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD
TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G (SEQ ID NO: 34)

Residue D10 of SEQ ID NO: 34, which is underlined and in bold, may be mutated *(e.g.,* to a A10) to yield a nuclease inactive SaCas9d.

Exemplary SaCas9n sequence

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD

TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G (SEQ ID NO: 35).

Residue A579 of SEQ ID NO: 35, which can be mutated from N579 of SEQ ID NO: 33 to yield a SaCas9 nickase, is underlined and in bold.

Exemplary SaKKH Cas9

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**AS**KKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD
TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNPvZLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFY^NDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPP***H***IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G (SEQ ID NO: 36).

Residue A579 of SEQ ID NO: 36, which can be mutated from N579 of SEQ ID NO: 36 to yield a SaCas9 nickase, is underlined and in bold. Residues K781, K967, and H1014 of SEQ

ID NO: 36, which can be mutated from E781, N967, and R1014 of SEQ ID NO: 36 to yield a SaKKH Cas9 are underlined and in italics.

[00280] In some embodiments, the Cas9 domain is a Cas9 domain from *Streptococcus pyogenes* (SpCas9). In some embodiments, the SpCas9 domain is a nuclease active SpCas9, a nuclease inactive SpCas9 (SpCas9d), or a SpCas9 nickase (SpCas9n). In some embodiments, the SpCas9 comprises the amino acid sequence SEQ ID NO: 37. In some embodiments, the SpCas9 comprises a D9X mutation of SEQ ID NO: 37, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, wherein X is any amino acid except for D. In some embodiments, the SpCas9 comprises a D9A mutation of SEQ ID NO: 37, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid seuqnce having a non-canonical PAM. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having a NGG, a NGA, or a NGCG PAM sequence. In some embodiments, the SpCas9 domain comprises one or more of a Dl 134X, a R1334X, and a T1336X mutation of SEQ ID NO: 37, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a Dl 134E, R1334Q, and T1336R mutation of SEQ ID NO: 37, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the SpCas9 domain comprises a D1134E, a R1334Q, and a T1336R mutation of SEQ ID NO: 37, or corresponding mutations in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the SpCas9 domain comprises one or more of a Dl 134X, a R1334X, and a T1336X mutation of SEQ ID NO: 37, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a Dl 134V, a R1334Q, and a T1336R mutation of SEQ ID NO: 37, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the SpCas9 domain comprises a Dl 134V, a R1334Q, and a T1336R mutation of SEQ ID NO: 37, or corresponding mutations in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the SpCas9 domain comprises one or more of a Dl 134X, a G1217X, a R1334X, and a T1336X mutation of SEQ ID NO: 37, or a

corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1134V, a G1217R, a R1334Q, and a T1336R mutation of SEQ ID NO: 37, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the SpCas9 domain comprises a D1134V, a G1217R, a R1334Q, and a T1336R mutation of SEQ ID NO: 37, or corresponding mutations in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein.

[00281] In some embodiments, the Cas9 domain of any of the fusion proteins provided herein comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of SEQ ID NOs: 37-41. In some embodiments, the Cas9 domain of any of the fusion proteins provided herein comprises the amino acid sequence of any one of SEQ ID NOs: 37-41. In some embodiments, the Cas9 domain of any of the fusion proteins provided herein consists of the amino acid sequence of any one of SEQ ID NOs: 37-41.

Exemplary SpCas9

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK YKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH

VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 37)

Exemplary SpCas9n

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK YKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 38)

Exemplary SpEQR Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEK YKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEE VVDKGASAQSFIERMTNFD KNLPNEKVLPKHS LLYEYFTVYNELTK V
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKS EQEIGKAT AKYFFYS NIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFESPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQ
YRSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 39)

Residues E1134, Q1334, and R1336 of SEQ ID NO: 39, which can be mutated from D1134, R1334, and T1336 of SEQ ID NO: 39 to yield a SpEQR Cas9, are underlined and in bold.

Exemplary SpVQR Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN

GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEK YKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEE VVDKGASAQSFIERMTNFD KNLPNEKVLPKHS LLYEYFTVYNELTK V
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKS EQEIGKAT AKYFFYS NIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQ
YRSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 40)

Residues VI 134, Q1334, and R1336 of SEQ ID NO: 40, which can be mutated from D1134, R1334, and T1336 of SEQ ID NO: 40 to yield a SpVQR Cas9, are underlined and in bold.

Exemplary SpVRER Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEK YKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEE VVDKGASAQSFIERMTNFD KNLPNEKVLPKHS LLYEYFTVYNELTK V
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED

RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKS EQEIGKAT AKYFFYS NIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAR
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKE
YRSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 41)

Residues VI 134, R1217, Q1334, and R1336 of SEQ ID NO: 41, which can be mutated from D1134, G1217, R1334, and T1336 of SEQ ID NO: 41 to yield a SpVRER Cas9, are underlined and in bold.

[00282] The following are exemplary fusion proteins *(e.g.,* base editing proteins) capable of binding to a nucleic acid sequence having a non-canonical *(e.g.,* a non-NGG) PAM sequence:

Exemplary SaBE3 (rAPOBECl-XTEN-SaCas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPESKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKR
GARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAAL
LHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRG
SINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWK
DIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEK
FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKE
IIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINL
ILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVIN
AIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIE
KIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEAS
KKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKD
FINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG
YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFI
TPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKD
NDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKY
SKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF

VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIG
VNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKS
KKHPQIIKKGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYD
ESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 42)

Exemplary SaKKH-BE3 (rAPOBECl-XTEN-SaCas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPESKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKR
GARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAAL
LHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRG
SINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWK
DIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEK
FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKE
IIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINL
ILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVIN
AIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIE
KIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEAS
KKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKD
FINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG
YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFI
TPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKD
NDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKY
SKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF
VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYENDLIKINGELYRVIG
VNNDLLNRIEVNMIDITYREYLENMNDKRPP*H*IIKTIASKTQSIKKYSTDILGNLYEVKS
KKHPQIIKKGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYD
ESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 43)

Exemplary EQR-BE3 (rAPOBECl-XTEN-Cas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE
DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH
FLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF

MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF*E*SPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKQYflSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS
NGENKIKMLSGGSPKKKRKV (SEQ ID NO: 44)

VQR-BE3 (rAPOBECl-XTEN-Cas9n-UGI-NLS)
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSггWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE
DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH
FLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFySPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKQYflSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS
NGENKIKMLSGGSPKKKRKV (SEQ ID NO: 45)

VRER-BE3 (rAPOBECl-XTEN-Cas9n-UGI-NLS)
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN

KHVEVNFIEKFTTERYFCPNTRCS IT WFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLIS SGVTIQIMTEQES GYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE
DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH
FLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFySPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAflELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRK*E*Y^STKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS
NGENKIKMLSGGSPKKKRKV (SEQ ID NO: 46)

High Fidelity Base Editors

[00283] Some aspects of the disclosure provide Cas9 fusion proteins *(e.g.,* any of the fusion proteins provided herein) comprising a Cas9 domain that has high fidelity. Additional aspects of the disclosure provide Cas9 fusion proteins *(e.g.,* any of the fusion proteins provided herein) comprising a Cas9 domain with decreased electrostatic interactions between the Cas9 domain and a sugar-phosphate backbone of a DNA, as compared to a wild-type Cas9 domain. In some embodiments, a Cas9 domain *(e.g.,* a wild type Cas9 domain) comprises one or more mutations that decreases the association between the Cas9 domain and a sugar-phosphate backbone of a DNA. In some embodiments, any of the Cas9 fusion proteins provided herein comprise one or more of a N497X, a R661X, a Q695X, and/or a Q926X mutation of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, wherein X is any amino acid. In some embodiments, any of the Cas9 fusion proteins provided herein comprise one or more of a N497A, a R661A, a Q695A, and/or a Q926A mutation of the amino acid sequence

provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the Cas9 domain comprises a D10A mutation of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the Cas9 domain *(e.g.,* of any of the fusion proteins provided herein) comprises the amino acid sequence as set forth in SEQ ID NO: 47. In some embodiments, the fusion protein comprises the amino acid sequence as set forth in SEQ ID NO: 48. Cas9 domains with high fidelity are known in the art and would be apparent to the skilled artisan. For example, Cas9 domains with high fidelity have been described in Kleinstiver, B.P., *et al.* "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." *Nature* 529, 490-495 (2016); and Slaymaker, I.M., *et al.* "Rationally engineered Cas9 nucleases with improved specificity." *Science* 351, 84-88 (2015); the entire contents of each are incorporated herein by reference.

[00284] It should be appreciated that the base editors provided herein, for example, base editor 2 (BE2) or base editor 3 (BE3), may be converted into high fidelity base editors by modifying the Cas9 domain as described herein to generate high fidelity base editors, for example, high fidelity base editor 2 (HF-BE2) or high fidelity base editor 3 (HF-BE3). In some embodiments, base editor 2 (BE2) comprises a deaminase domain, a dCas9, and a UGI domain. In some embodiments, base editor 3 (BE3) comprises a deaminase domain, anCas9 domain and a UGI domain.

Cas9 domain where mutations relative to Cas9 of SEQ ID NO: 6 are shown in bold and underlines

DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAI
LSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK
QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSL
TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG

LSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQ
KQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 47)

HF-BE3

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFT
TERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQI
MTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPN
FKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK
DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV
KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY
NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GD (SEQ ID NO: 48)

Cas9 fusion proteins

[00285] Any of the Cas9 domains *{e.g.*, a nuclease active Cas9 protein, a nuclease-inactive dCas9 protein, or a Cas9 nickase protein) disclosed herein may be fused to a second protein, thus fusion proteins provided herein comprise a Cas9 domain as provided herein and a second protein, or a "fusion partner". In some embodiments, the second protein is fused to the N-

terminus of the Cas9 domain. However, in other embodiments, the second protein is fused to the C-terminus of the Cas9 domain. In some embodiments, the second protein that is fused to the Cas9 domain is a nucleic acid editing domain. In some embodiments, the Cas9 domain and the nucleic acid editing domain are fused via a linker, while in other embodiments the Cas9 domain and the nucleic acid editing domain are fused directly to one another. In some embodiments, the Cas9 domain and the nucleic acid editing domain are fused via a linker of any length or composition. For example, the linker may be a bond, one or more amino acids, a peptide, or a polymer, of any length and composition. In some embodiments, the linker comprises $(GGGS)_n$ (SEQ ID NO: 613), $(GGGGS)_n$ (SEQ ID NO: 607), $(G)_n$ (SEQ ID NO: 608), $(EAAAK)_n$ (SEQ ID NO: 609), $(GGS)_n$ (SEQ ID NO: 610), $(SGGS)_n$ (SEQ ID NO: 606), SGSETPGTSESATPES (SEQ ID NO: 604), $SGGS(GGS)_n$ (SEQ ID NO: 612), SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 605), or $(XP)_n$ (SEQ ID NO: 611) motif, or a combination of any of these, wherein n is independently an integer between 1 and 30, and wherein X is any amino acid. In some embodiments, the linker comprises a $(GGS)_n$ motif, wherein n is 1, 3, or 7. In some embodiments, the linker comprises a $(GGS)_n$ (SEQ ID NO: 610) motif, wherein n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. In some embodiments, the linker comprises the amino acid sequence $SGGS(GGS)_n$ (SEQ ID NO: 612), wherein n is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. In some embodiments, the linker comprises the amino acid sequence $SGGS(GGS)_n$ (SEQ ID NO: 612), wherein n is 2. In some embodiments, the linker comprises an amino acid sequence of SGSETPGTSESATPES (SEQ ID NO: 604), also referred to as the XTEN linker in the Examples). In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 605), also referred to as the 32 amino acid linker in the Examples. The length of the linker can influence the base to be edited, as illustrated in the Examples. For example, a linker of 3-amino-acid long *(e.g.,* **(GGS)i)** may give a 2-5, 2-4, 2-3, 3-4 base editing window relative to the PAM sequence, while a 9-amino-acid linker *(e.g.,* $(GGS)_3$ (SEQ ID NO: 610)) may give a 2-6, 2-5, 2-4, 2-3, 3-6, 3-5, 3-4, 4-6, 4-5, 5-6 base editing window relative to the PAM sequence. A 16-amino-acid linker *(e.g.,* the XTEN linker) may give a 2-7, 2-6, 2-5, 2-4, 2-3, 3-7, 3-6, 3-5, 3-4, 4-7, 4-6, 4-5, 5-7, 5-6, 6-7 base window relative to the PAM sequence with exceptionally strong activity, and a 21-amino-acid linker *(e.g.,* $(GGS)_7$ (SEQ ID NO: 610)) may give a 3-8, 3-7, 3-6, 3-5, 3-4, 4-8, 4-7, 4-6, 4-5, 5-8, 5-7, 5-6, 6-8, 6-7, 7-8 base editing window relative to the PAM sequence. The novel finding that varying linker length may allow the dCas9 fusion proteins of the disclosure to edit nucleobases different distances from the PAM sequence affords siginicant clinical importance, since a PAM sequence may be of

varying distance to the disease-causing mutation to be corrected in a gene. It is to be understood that the linker lengths described as examples here are not meant to be limiting.

[00286] In some embodiments, the second protein comprises an enzymatic domain. In some embodiments, the enzymatic domain is a nucleic acid editing domain. Such a nucleic acid editing domain may be, without limitation, a nuclease, a nickase, a recombinase, a deaminase, a methyltransferase, a methylase, an acetylase, or an acetyltransferase. Non-limiting exemplary binding domains that may be used in accordance with this disclosure include transcriptional activator domains and transcriptional repressor domains.

Deaminase Domains

[00287] In some embodiments, second protein comprises a nucleic acid editing domain. In some embodiments, the nucleic acid editing domain can catalyze a C to U base change. In some embodiments, the nucleic acid editing domain is a deaminase domain. In some embodiments, the deaminase is a cytidine deaminase or a cytidine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBECl deaminase. In some embodiments, the deaminase is an APOBEC2 deaminase. In some embodiments, the deaminase is an APOBEC3 deaminase. In some embodiments, the deaminase is an APOBEC3A deaminase. In some embodiments, the deaminase is an APOBEC3B deaminase. In some embodiments, the deaminase is an APOBEC3C deaminase. In some embodiments, the deaminase is an APOBEC3D deaminase. In some embodiments, the deaminase is an APOBEC3E deaminase. In some embodiments, the deaminase is an APOBEC3F deaminase. In some embodiments, the deaminase is an APOBEC3G deaminase. In some embodiments, the deaminase is an APOBEC3H deaminase. In some embodiments, the deaminase is an APOBEC4 deaminase. In some embodiments, the deaminase is an activation-induced deaminase (AID). In some embodiments, the deaminase is a vertebrate deaminase. In some embodiments, the deaminase is an invertebrate deaminase. In some embodiments, the deaminase is a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse deaminase. In some embodiments, the deaminase is a human deaminase. In some embodiments, the deaminase is a rat deaminase, *e.g.*, rAPOBECl. In some embodiments, the deaminase is an activation-induced cytidine deaminase (AID). In some embodiments, the deaminase is a cytidine deaminase 1 (CDA1). In some embodiments, the deaminase is a *Petromyzon marinus* cytidine deaminase 1 (pmCDAl). In some embodiments, the deminase is a human APOBEC3G (SEQ ID NO: 60). In some embodiments, the deaminase is a fragment of the

human APOBEC3G (SEQ ID NO: 83). In some embodiments, the deaminase is a human APOBEC3G variant comprising a D316R_D317R mutation (SEQ ID NO: 82). In some embodiments, the deaminase is a frantment of the human APOBEC3G and comprising mutations corresponding to the D316R_D317R mutations in SEQ ID NO: 60 (SEQ ID NO: 84).

[00288] In some embodiments, the nucleic acid editing domain is at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the deaminase domain of any one of SEQ ID NOs: 49-84. In some embodiments, the nucleic acid editing domain comprises the amino acid sequence of any one of SEQ ID NOs: 49-84.

Deaminase Domains that Modulate the Editing Window of Base Editors

[00289] Some aspects of the disclosure are based on the recognition that modulating the deaminase domain catalytic activity of any of the fusion proteins provided herein, for example by making point mutations in the deaminase domain, affect the processivity of the fusion proteins *{e.g.,* base editors). For example, mutations that reduce, but do not eliminate, the catalytic activity of a deaminase domain within a base editing fusion protein can make it less likely that the deaminase domain will catalyze the deamination of a residue adjacent to a target residue, thereby narrowing the deamination window. The ability to narrow the deaminataion window may prevent unwanted deamination of residues adjacent of specific target residues, which may decrease or prevent off-target effects.

[00290] In some embodiments, any of the fusion proteins provided herein comprise a deaminase domain *{e.g.,* a cytidine deaminase domain) that has reduced catalytic deaminase activity. In some embodiments, any of the fusion proteins provided herein comprise a deaminase domain *{e.g.,* a cytidine deaminase domain) that has a reduced catalytic deaminase activity as compared to an appropriate control. For example, the appropriate control may be the deaminase activity of the deaminase prior to introducing one or more mutations into the deaminase. In other embodiments, the appropriate control may be a wild-type deaminase. In some embodiments, the appropriate control is a wild-type apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the appropriate control is an APOBEC1 deaminase, an APOBEC2 deaminase, an APOBEC3A deaminase, an APOBEC3B deaminase, an APOBEC3C deaminase, an APOBEC3D deaminase, an APOBEC3F deaminase, an APOBEC3G deaminase, or an APOBEC3H deaminase. In some embodiments, the appropriate control is an activation induced deaminase (AID). In some embodiments, the

appropriate control is a cytidine deaminase 1 from *Petromyzon marinm* (pmCDAl). In some embodiments, the deaminse domain may be a deaminase domain that has at least 1%, at least 5%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at lest 70%, at least 80%, at least 90%, or at least 95% less catalytic deaminase activity as compared to an appropriate control.

[00291] In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising one or more mutations selected from the group consisting of H121X, H122X, R126X, R126X, R118X, W90X, W90X, and R132X of rAPOBECl (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase, wherin X is any amino acid. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising one or more mutations selected from the group consisting of H121R, H122R, R126A, R126E, R118A, W90A, W90Y, and R132E of rAPOBECl (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase.

[00292] In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising one or more mutations selected from the group consisting of D316X, D317X, R320X, R320X, R313X, W285X, W285X, R326X of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase, wherin X is any amino acid. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising one or more mutations selected from the group consisting of D316R, D317R, R320A, R320E, R313A, W285A, W285Y, R326E of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase.

[00293] In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a H121R and a H122Rmutation of rAPOBECl (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R126A mutation of rAPOBECl (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R126E mutation of rAPOBECl (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R118A mutation of rAPOBECl (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a W90A

mutation of rAPOBEC1 (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a W90Y mutation of rAPOBEC1 (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R132E mutation of rAPOBEC1 (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a W90Y and a R126E mutation of rAPOBEC1 (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R126E and a R132E mutation of rAPOBEC1 (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a W90Y and a R132E mutation of rAPOBEC1 (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a W90Y, R126E, and R132E mutation of rAPOBEC1 (SEQ ID NO: 76), or one or more corresponding mutations in another APOBEC deaminase.

[00294] In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a D316R and a D317R mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R320A mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R320E mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R313A mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a W285A mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a W285Y mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in

another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R326E mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a W285Y and a R320E mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R320E and a R326E mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a W285Y and a R326E mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a W285Y, R320E, and R326E mutation of hAPOBEC3G (SEQ ID NO: 60), or one or more corresponding mutations in another APOBEC deaminase.

[00295] Some aspects of this disclosure provide fusion proteins comprising (i) a nuclease-inactive Cas9 domain; and (ii) a nucleic acid editing domain. In some embodiments, a nuclease-inactive Cas9 domain (dCas9), comprises an amino acid sequence that is at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 as provided by any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, and comprises mutations that inactivate the nuclease activity of Cas9. Mutations that render the nuclease domains of Cas9 inactive are well-known in the art. For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvCl subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvCl subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations DIOA and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al., Science.* 337:816-821(2012); Qi *et al., Cell.* 28;152(5): 1173-83 (2013)). In some embodiments, the dCas9 of this disclosure comprises a DIOA mutation of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the dCas9 of this disclosure comprises a H840A mutation of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as

any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the dCas9 of this disclosure comprises both D10A and H840A mutations of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the Cas9 further comprises a histidine residue at position 840 of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. The presence of the catalytic residue H840 restores the acvitity of the Cas9 to cleave the non-edited strand containing a G opposite the targeted C. Restoration of H840 does not result in the cleavage of the target strand containing the C. In some embodiments, the dCas9 comprises an amino acid sequence of SEQ ID NO: 32. It is to be understood that other mutations that inactivate the nuclease domains of Cas9 may also be included in the dCas9 of this disclosure.

[00296] The Cas9 or dCas9 domains comprising the mutations disclosed herein, may be a full-length Cas9, or a fragment thereof. In some embodiments, proteins comprising Cas9, or fragments thereof, are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to wild type Cas9. In some embodiments, the Cas9 variant comprises a fragment of Cas9 *(e.g.,* a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas9, *e.g.,* a Cas9 comprising the amino acid sequence of SEQ ID NO: 6.

[00297] Any of the Cas9 fusion proteins of this disclosure may further comprise a nucleic acid editing domain *(e.g.,* an enzyme that is capable of modifying nucleic acid, such as a deaminase). In some embodiments, the nucleic acid editing domain is a DNA-editing domain. In some embodiments, the nucleic acid editing domain has deaminase activity. In some embodiments, the nucleic acid editing domain comprises or consists of a deaminase or deaminase domain. In some embodiments, the deaminase is a cytidine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBEC 1 family deaminase.

In some embodiments, the deaminase is an activation-induced cytidine deaminase (AID). Some nucleic-acid editing domains as well as Cas9 fusion proteins including such domains are described in detail herein. Additional suitable nucleic acid editing domains will be apparent to the skilled artisan based on this disclosure and knowledge in the field.

[00298] Some aspects of the disclosure provide a fusion protein comprising a Cas9 domain fused to a nucleic acid editing domain, wherein the nucleic acid editing domain is fused to the N-terminus of the Cas9 domain. In some embodiments, the Cas9 domain and the nucleic acid editing-editing domain are fused via a linker. In some embodiments, the linker comprises a $(GGGS)_n$ (SEQ ID NO: 613), a $(GGGGS)_n$ (SEQ ID NO: 607), a $(G)_n$ (SEQ ID NO: 608), an $(EAAAK)_n$ (SEQ ID NO: 609), a $(GGS)_n$ (SEQ ID NO: 610), $(SGGS)_n$ (SEQ ID NO: 606), an SGSETPGTSESATPES (SEQ ID NO: 604) motif (see, *e.g.,* Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to Fokl nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014; 32(6): 577-82; the entire contents are incorporated herein by reference), or an $(XP)_n$ (SEQ ID NO: 611) motif, or a combination of any of these, wherein n is independently an integer between 1 and 30. In some embodiments, n is independently 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30, or, if more than one linker or more than one linker motif is present, any combination thereof. In some embodiments, the linker comprises a $(GGS)_n$ (SEQ ID NO: 610) motif, wherein n is 1,2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15. In some embodiments, the linker comprises a $(GGS)_n$ (SEQ ID NO: 610) motif, wherein n is 1, 3, or 7. In some embodiments, the linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 604). Additional suitable linker motifs and linker configurations will be apparent to those of skill in the art. In some embodiments, suitable linker motifs and configurations include those described in Chen *et al.,* Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev.* 2013; 65(10): 1357-69, the entire contents of which are incorporated herein by reference. Additional suitable linker sequences will be apparent to those of skill in the art based on the instant disclosure. In some embodiments, the general architecture of exemplary Cas9 fusion proteins provided herein comprises the structure:

$[NH_2]$-[nucleic acid editing domain]-[Cas9]-[COOH] or

$[NH_2]$-[nucleic acid editing domain]-[linker]-[Cas9]-[COOH],

wherein NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein.

[00299] The fusion proteins of the present disclosure may comprise one or more additional features. For example, in some embodiments, the fusion protein comprises a nuclear

localization sequence (NLS). In some embodiments, the NLS of the fusion protein is localized between the nucleic acid editing domain and the Cas9 domain. In some embodiments, the NLS of the fusion protein is localized C-terminal to the Cas9 domain.

[00300] Other exemplary features that may be present are localization sequences, such as cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags *(e.g.,* Softag 1, Softag 3), strep-tags , biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. In some embodiments, the fusion protein comprises one or more His tags.

[00301] In some embodiments, the nucleic acid editing domain is a deaminase. For example, in some embodiments, the general architecture of exemplary Cas9 fusion proteins with a deaminase domain comprises the structure:

$[NH_2]$-[NLS]-[deaminase]-[Cas9]-[COOH],

$[NH_2]$-[Cas9]-[deaminase]-[COOH],

$[NH_2]$-[deaminase]-[Cas9]-[COOH], or

$[NH_2]$-[deaminase]-[Cas9]-[NLS]-[COOH];

wherein NLS is a nuclear localization sequence, NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. Nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.,* PCT/EP2000/01 1690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, a NLS comprises the amino acid sequence PKKKRKV (SEQ ID NO: 614) or MDSLLMNRRKFLYQFKNVRW AKGRRET YLC (SEQ ID NO: 615). In some embodiments, a linker is inserted between the Cas9 and the deaminase. In some embodiments, the NLS is located C-terminal of the Cas9 domain. In some embodiments, the NLS is located N-terminal of the Cas9 domain. In some embodiments, the NLS is located between the deaminase and the Cas9 domain. In some embodiments, the NLS is located N-terminal of the deaminase domain. In some embodiments, the NLS is located C-terminal of the deaminase domain.

[00302] One exemplary suitable type of nucleic acid editing domain is a cytidine deaminase, for example, of the APOBEC family. The apolipoprotein B mRNA-editing complex (APOBEC) family of cytidine deaminase enzymes encompasses eleven proteins that serve to initiate mutagenesis in a controlled and beneficial manner.[29] One family member, activation-induced cytidine deaminase (AID), is responsible for the maturation of antibodies by converting cytosines in ssDNA to uracils in a transcription-dependent, strand-biased fashion.[30] The apolipoprotein B editing complex 3 (APOBEC3) enzyme provides protection to human cells against a certain HIV-1 strain via the deamination of cytosines in reverse-transcribed viral ssDNA.[31] These proteins all require a Zn^{2+}-coordinating motif (His-X-Glu-X23_26-Pro-Cys-$X2_{-4}$-Cys; SEQ ID NO: 616) and bound water molecule for catalytic activity. The Glu residue acts to activate the water molecule to a zinc hydroxide for nucleophilic attack in the deamination reaction. Each family member preferentially deaminates at its own particular "hotspot", ranging from WRC (W is A or T, R is A or G) for hAID, to TTC for hAPOBEC3F.[32] A recent crystal structure of the catalytic domain of APOBEC3G revealed a secondary structure comprised of a five-stranded β-sheet core flanked by six a-helices, which is believed to be conserved across the entire family.[33] The active center loops have been shown to be responsible for both ssDNA binding and in determining "hotspot" identity.[34] Overexpression of these enzymes has been linked to genomic instability and cancer, thus highlighting the importance of sequence-specific targeting.[35]

[00303] Some aspects of this disclosure relate to the recognition that the activity of cytidine deaminase enzymes such as APOBEC enzymes can be directed to a specific site in genomic DNA. Without wishing to be bound by any particular theory, advantages of using Cas9 as a recognition agent include (1) the sequence specificity of Cas9 can be easily altered by simply changing the sgRNA sequence; and (2) Cas9 binds to its target sequence by denaturing the dsDNA, resulting in a stretch of DNA that is single-stranded and therefore a viable substrate for the deaminase. It should be understood that other catalytic domains, or catalytic domains from other deaminases, can also be used to generate fusion proteins with Cas9, and that the disclosure is not limited in this regard.

[00304] Some aspects of this disclosure are based on the recognition that Cas9:deaminase fusion proteins can efficiently deaminate nucleotides at positions 3-11 according to the numbering scheme in Figure 3. In view of the results provided herein regarding the nucleotides that can be targeted by Cas9:deaminase fusion proteins, a person of skill in the art will be able to design suitable guide RNAs to target the fusion proteins to a target sequence that comprises a nucleotide to be deaminated.

[00305] In some embodiments, the deaminase domain and the Cas9 domain are fused to each other via a linker. Various linker lengths and flexibilities between the deaminase domain *(e.g.,* AID) and the Cas9 domain can be employed *(e.g.,* ranging from very flexible linkers of the form (GGGGS)„ (SEQ ID NO: 607), (GGS)„(SEQ ID NO: 610), and (G)„ (SEQ ID NO: 608) to more rigid linkers of the form $(EAAAK)_n$ (SEQ ID NO: 609), $(SGGS)_n$ (SEQ ID NO: 606), SGSETPGTSESATPES (SEQ ID NO: 604) (see, *e.g.,* Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to Fokl nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014; 32(6): 577-82; the entire contents are incorporated herein by reference) and $(XP)_n$ (SEQ ID NO: 611))[36] in order to achieve the optimal length for deaminase activity for the specific application. In some embodiments, the linker comprises a $(GGS)_n$ (SEQ ID NO: 610) motif, wherein n is 1, 3, or 7. In some embodiments, the linker comprises a (an SGSETPGTSESATPES (SEQ ID NO: 604) motif.

[00306] Some exemplary suitable nucleic-acid editing domains, *e.g.,* deaminases and deaminase domains, that can be fused to Cas9 domains according to aspects of this disclosure are provided below. It should be understood that, in some embodiments, the active domain of the respective sequence can be used, *e.g.,* the domain without a localizing signal (nuclear localization sequence, without nuclear export signal, cytoplasmic localizing signal).

[00307] Human AID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS
VRLSROLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 49)

(underline: nuclear localization sequence; double underline: nuclear export signal)

[00308] Mouse AID:

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENS
VRLTROLRRILLPLYEVDDLRDAFRMLGF (SEQ ID NO: 51)

(underline: nuclear localization sequence; double underline: nuclear export signal)

[00309] Dog AID:

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAAR

LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENS VRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 52)

(underline: nuclear localization sequence; double underline: nuclear export signal)

[00310] Bovine AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCH VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARL YFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS VRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 53)

(underline: nuclear localization sequence; double underline: nuclear export signal)

[00311] Rat:AID:MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRW LRPAATQDPVSPPRSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLD FGYLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGN PNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTFV ENHERTFKAWEGLHENSV RLSRRLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 54)

(underline: nuclear localization sequence; double underline: nuclear export signal)

[00312] Mouse APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVS LHH*GV*FKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAE*QI*V*RFL*AT HHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGR RFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMD PLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ *HAE/* ***LFLDKIRSMELSQWITCYLTWSPCPNCAWQLAAFKRORPOLlLmYTSRLYFFiWKRPFQ*** KGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIIS RRTQRRLRRIKES WGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 55)

(italic: nucleic acid editing domain)

[00313] Rat APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVSL *miGVFKNKONmAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLAT* HHNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGR RFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRRVH LLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQ *HAE/* *LFLDKIRSMELSQVIITCYLTWSPCPNCAWQLAAFKRORPOULmYTSRLYFFiWKRPFQ* KGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIIS RRTQRRLHRIKES WGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 56)

(italic: nucleic acid editing domain)

[00314] Rhesus macaque APOBEC-3G:

MVEPMDPRTFVSNFNNRPILS GLNTVWLCCEVKTKDPS GPPLDAKIFQGKVYS KAKY *HPEMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKOPKVTLTIFVAR* LYYFWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNN LPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWV ***PLNQHRGFLRNQAPNinGFPKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSCAQ*** EMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTF VDRQGRPFQPWDGLDEHS QALS GRLRAI (SEQ ID NO: 57)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[00315] Chimpanzee APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQ VYSKLKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTROyATFLAEOPKVT* LTIFVARLYYFWDPDYQEALRS LCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELF EPWNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERLHND *TWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLHQDYRVTCFTSWSPC FS*CAQEMAKFIS NNKHVSLCIFAARIYDDQGRC QEGLRTLAKAG AKIS IMTYSEFKHC WDTFVDHQGCPFQPWDGLEEHS QALS GRLRAILQNQGN (SEQ ID NO: 58)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[00316] Green monkey APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKL *YPEAKDHPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRCANSVATFLAEDP*KVTL TIFVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPF KPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSH *NOTWVLLNQHRGFLRNQAPORHGFPKGRHAELCFLDLIPFWKLDDQQYRVTCFTSWSP* CFSCAQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFE YCWDTFVDRQGRPFQPWDGLDEHS QALS GRLRAI (SEQ ID NO: 59)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[00317] Human APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQ *VYS ELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVT* LTIFVARLYYFWDPDYQEALRS LCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELF EPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHN *DTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSP*

CFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHC
WDTFVDHQGCPFQPWDGLDEHS QDLSGRLRAILQNQEN (SEQ ID NO: 60)
(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[00318] Human APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQ
VYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAERPNVTL
TISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWY
KFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKH
HSPVSWKRGVERNQVOPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGE
VAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENF
VYNDDEPFKPWKGLKYNFLFLDS KLQEILE (SEQ ID NO: 61)
(italic: nucleic acid editing domain)

[00319] Human APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRG
QVYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSERPNVT
LTISAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPW
YKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTW
VLMOQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSW
GCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYC
WDTFVYRQGCPFQPWDGLEEHS QALSGRLRAILQNQGN (SEQ ID NO: 62)
(italic: nucleic acid editing domain)

[00320] Rat APOBEC-3B:

MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNN
FLCYEVNGMDCALPVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTW
YMSWSPCSKCAEQVARFLAAHRNLSLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVA
AMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLREDVFYLQ
FNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMR
SMELSQVRITCYLTWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFYWRKKFQKGLCT
LWRSGIHVDVMDLPQFADCWTNFVNPQRPFRPWNELEKNSWRIQRRLRRIKESWGL
(SEQ ID NO: 63)

[00321] Bovine APOBEC-3B:

DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLF
KQQFGNQPRVPAPYYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAEIRFIDKINSL
DLNPSQSYKIICYITWSPCPNCANELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQD

LQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQRILTAPI (SEQ ID NO: 64)

[00322] Chimpanzee APOBEC-3B:

MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFR GQMYSQPEHHAEMCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHP NVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQPF MPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEVERLDN GTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISW SPCFSWGCAGQVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTY DEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPP QSPGPCLPLCSEPPLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPVPSF HSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG (SEQ ID NO: 65)

[00323] Human APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR *NQVOSETHCHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVN* LTIFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWK GLKTNFRLLKRRLRESLQ (SEQ ID NO: 66)

(italic: nucleic acid editing domain)

[00324] Gorilla APOBEC3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR *NQVOSETHCHAERCFLSWFCDDILSPNTNYQVTWYTSWSPCPECAGEVAEFLARHSNVN* LTIFTARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPW KGLKYNFRFLKRRLQEILE (SEQ ID NO: 67)

(italic: nucleic acid editing domain)

[00325] Human APOBEC-3A:

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHN *QAKNLLCGEYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQE* NTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCP FQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 68)

(italic: nucleic acid editing domain)

[00326] Rhesus macaque APOBEC-3A:

MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRG *FLCNKAKNVPCGOYGCHVELRFLCEVPSWQLDPAQTYRVTWFISWSPCFRRGCAGQVR*

VFLQENKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDR QGRPFQPWDGLDEHS QALS GRLRAILQNQGN (SEQ ID NO: 69)

(italic: nucleic acid editing domain)

[00327] Bovine APOBEC-3A:

MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC *H AELYFLGKIHS WNLDRNQHYRLTCFIS* W*SPC* YDCAQKLTTFLKENHHIS LHILAS RIYTHN RFGCHQS GLCELQAAGARITIMTFEDFKHC WETFVDHKGKPFQPWEGLNVKS QALCT ELQAILKTQQN (SEQ ID NO: 70)

(italic: nucleic acid editing domain)

[00328] Human APOBEC-3H:

MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC *HAE*/C *FINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVOF1KAHOHLNLG1FASRLYYHWCK* PQQKGLRLLC GSQVPVEVMGFPKF ADC WENF VDHEKPLS FNP YKMLEELD KNS RAIK RRLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO: 71)

(italic: nucleic acid editing domain)

[00329] Rhesus macaque APOBEC-3H:

MALLT AKTFS LQFNNKRRVNKP YYPRKALLC YQLTPQNGS TPTRGHLKNKKKDH AEI RFINKIKS MGLDETQC YQVTC YLTWS PCPS CAGELVDFIKAHRHLNLRIF A SRLY YHW RPNYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAI KRRLERIKS RSVDVLENGLRS LQLGP VTPS SSIRNS R (SEQ ID NO: 72)

[00330] Human APOBEC-3D:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRG *PVLPKRQSNHRQEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVK* VTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWE NFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRN ESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHC *HAERCFLSWFCDD/LSPNTNYEVT WYTSWSPCPE*CAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKI MGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ (SEQ ID NO: 73)

(italic: nucleic acid editing domain)

[00331] Human APOBEC- 1:

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNT TNH VEVNFIKKFTS ERDFHPS MSC SITWFLS W SPCWECS QAIREFLS RHPG VTLVIY V A RLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPP

LWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPS VAWR (SEQ ID NO: 74)

[00332] Mouse APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNT SNHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIAR LYHHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWV KLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK (SEQ ID NO: 75)

[00333] Rat APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 76)

[00334] Human APOBEC-2:

MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRN VE

YSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRY NVTWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCK LRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK (SEQ ID NO: 77)

[00335] Mouse APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRN VEYSSGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEE AFFNTILPAFDPAL KYNVTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEA GCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK (SEQ ID NO: 78)

[00336] Rat APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRN VEYSSGRNKTFLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEE AFFNTILPAFDPAL KYNVTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEA GCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK (SEQ ID NO: 79)

[00337] Bovine APOBEC-2:

MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRN
VE
YSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRY
MVTWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGC
RLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK (SEQ ID
NO: 80)

[00338] Petromyzon marinus CDA1 (pmCDAl)
MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKP
QSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGN
GHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQL
NENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO: 81)

[00339] Human APOBEC3G D316R_D317R
MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQ
VYSELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPK
VTLTIFVARLYYFWDPD YQEALRSLCQKRDGPRATMKIMN YDEFQHCWS KFVYS QR
ELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERM
HNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFT
SWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISIMTYSE
FKHCWDTFVDHQGCPFQPWDGLDEHS QDLSGRLRAILQNQEN (SEQ ID NO: 82)

[00340] Human APOBEC3G chain A
MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGF
LEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIF
TARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEH
SQDLSGRLRAILQ (SEQ ID NO: 83)

[00341] Human APOBEC3G chain A D120R_D121R
MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGF
LEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIF
TARIYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEH
SQDLSGRLRAILQ (SEQ ID NO: 84)

[00342] In some embodiments, fusion proteins as provided herein comprise the full-length amino acid of a nucleic acid editing enzyme, *e.g.,* one of the sequences provided above. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length sequence of a nucleic acid editing enzyme, but only a fragment thereof. For example, in some

embodiments, a fusion protein provided herein comprises a Cas9 domain and a fragment of a nucleic acid editing enzyme, *e.g.*, wherein the fragment comprises a nucleic acid editing domain. Exemplary amino acid sequences of nucleic acid editing domains are shown in the sequences above as italicized letters, and additional suitable sequences of such domains will be apparent to those of skill in the art.

[00343] Additional suitable nucleic-acid editing enzyme sequences, *e.g.*, deaminase enzyme and domain sequences, that can be used according to aspects of this invention, *e.g.*, that can be fused to a nuclease-inactive Cas9 domain, will be apparent to those of skill in the art based on this disclosure. In some embodiments, such additional enzyme sequences include deaminase enzyme or deaminase domain sequences that are at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% similar to the sequences provided herein. Additional suitable Cas9 domains, variants, and sequences will also be apparent to those of skill in the art. Examples of such additional suitable Cas9 domains include, but are not limited to, D10A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (see, *e.g.*, Prashant *et al*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature Biotechnology. 2013; 31(9): 833-838 the entire contents of which are incorporated herein by reference). In some embodiments, the Cas9 comprises a histidine residue at position 840 of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. The presence of the catalytic residue H840 restores the acvitity of the Cas9 to cleave the non-edited strand containing a G opposite the targeted C. Restoration of H840 does not result in the cleavage of the target strand containing the C.

[00344] Additional suitable strategies for generating fusion proteins comprising a Cas9 domain and a deaminase domain will be apparent to those of skill in the art based on this disclosure in combination with the general knowledge in the art. Suitable strategies for generating fusion proteins according to aspects of this disclosure using linkers or without the use of linkers will also be apparent to those of skill in the art in view of the instant disclosure and the knowledge in the art. For example, Gilbert *et al.*, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* **2013;** 154(2):442-51, showed that C-terminal fusions of Cas9 with VP64 using 2 NLS's as a linker (SPKKKRKVEAS, SEQ ID NO: 617), can be employed for transcriptional activation. Mali *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* **2013;** 31(9):833-8, reported that C-terminal fusions with VP64

without linker can be employed for transcriptional activation. And Maeder *et al.*, CRISPR RNA-guided activation of endogenous human genes. *Nat Methods.* **2013;** 10: 977-979, reported that C-terminal fusions with VP64 using a Gly_4Ser (SEQ ID NO: 613) linker can be used as transcriptional activators. Recently, dCas9- Fokl nuclease fusions have successfully been generated and exhibit improved enzymatic specificity as compared to the parental Cas9 enzyme (In Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to Fokl nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014; 32(6): 577-82, and in Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. Dimeric CRISPR RNA-guided Fokl nucleases for highly specific genome editing. *Nat Biotechnol.* 2014; 32(6):569-76. PMJX>: 24770325 a SGSETPGTSESATPES (SEQ ID NO: 604) or a GGGGS (SEQ ID NO: 607) linker was used in FokI-dCas9 fusion proteins, respectively).

[00345] Some aspects of this disclosure provide fusion proteins comprising (i) a Cas9 enzyme or domain (e.g., a first protein); and (ii) a nucleic acid-editing enzyme or domain (e.g., a second protein). In some aspects, the fusion proteins provided herein further include (iii) a programmable DNA-binding protein, for example, a zinc-finger domain, a TALE, or a second Cas9 protein (e.g., a third protein). Without wishing to be bound by any particular theory, fusing a programmable DNA-binding protein (e.g., a second Cas9 protein) to a fusion protein comprising (i) a Cas9 enzyme or domain (e.g., a first protein); and (ii) a nucleic acid-editing enzyme or domain (e.g., a second protein) may be useful for improving specificity of the fusion protein to a target nucleic acid sequence, or for improving specificity or binding affinity of the fusion protein to bind target nucleic acid sequence that does not contain a canonical PAM (NGG) sequence. In some embodiments, the third protein is a Cas9 protein (e.g, a second Cas9 protein). In some embodiments, the third protein is any of the Cas9 proteins provided herein. In some embodiments, the third protein is fused to the fusion protein N-terminal to the Cas9 protein (e.g., the first protein). In some embodiments, the third protein is fused to the fusion protein C-terminal to the Cas9 protein (e.g., the first protein). In some embodiments, the Cas9 domain (e.g., the first protein) and the third protein (e.g., a second Cas9 protein) are fused via a linker (e.g., a second linker). In some embodiments, the linker comprises a $(GGGGS)_n$ (SEQ ID NO: 607), a $(G)_n$ (SEQ ID NO: 608), an $(EAAAK)_n$ (SEQ ID NO: 609), a $(GGS)_n$ (SEQ ID NO: 610), $(SGGS)_n$ (SEQ ID NO: 606), a SGSETPGTSESATPES (SEQ ID NO: 604), a $SGGS(GGS)_n$ (SEQ ID NO: 612), a SGGSSGGSSGS ETPGTS ESATPES SGGSSGGS (SEQ ID NO: 605), or an $(XP)_n$ (SEQ ID NO: 611) motif, or a combination of any of these, wherein n is independently an integer

between 1 and 30. In some embodiments, the general architecture of exemplary napDNAbp fusion proteins provided herein comprises the structure:

[NH2]-[nucleic acid-editing enzyme or domain]-[napDNAbp]-[third protein]-[COOH];

[NH2]-[third protein]-[napDNAbp]-[nucleic acid-editing enzyme or domain]-[COOH];

[NH2]-[napDNAbp]-[nucleic acid-editing enzyme or domain]-[third protein]-[COOH];

[NH2]-[third protein]-[nucleic acid-editing enzyme or domain]-[napDNAbp]-[COOH];

[NH2]-[UGI]-[nucleic acid-editing enzyme or domain]- napDNAbp]-[third protein]-[COOH];

[NH2]-[UGI]-[third protein]-[napDNAbp]-[nucleic acid-editing enzyme or domain]-[COOH];

[NH2]-[UGI]-[napDNAbp]-[nucleic acid-editing enzyme or domain]-[third protein]-[COOH];

[NH2]-[UGI]-[third protein]-[nucleic acid-editing enzyme or domain]-[napDNAbp]-[COOH];

[NH2]-[nucleic acid-editing enzyme or domain]-[napDNAbp]-[third protein]-[UGI]-[COOH];

[NH2]-[third protein]-[napDNAbp]-[nucleic acid-editing enzyme or domain]-[UGI]-[COOH];

[NH2]-[NapDNAbp]-[nucleic acid-editing enzyme or domain]-[third protein]-[UGI]-[COOH];

[NH2]-[third protein]-[nucleic acid-editing enzyme or domain]-[NapDNAbp]-[UGI]-[COOH];

or

[NH2]-[nucleic acid-editing enzyme or domain]-[NapDNAbp]-[first UGI domain]-[second UGI domain]-[COOH];

wherein NH2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, the "]-[" used in the general architecture above indicates the presence of an optional linker sequence. In other examples, the general architecture of exemplary NapDNAbp fusion proteins provided herein comprises the structure:

[NH2]-[nucleic acid-editing enzyme or domain]-[NapDNAbp]-[second NapDNAbp protein]-[COOH];

[NH2]-[second NapDNAbp protein]-[NapDNAbp]-[nucleic acid-editing enzyme or domain]-[COOH];

[NH2]-[NapDNAbp]-[nucleic acid-editing enzyme or domain]-[second NapDNAbp protein]-[COOH];

[NH2]-[second NapDNAbp protein]-[nucleic acid-editing enzyme or domain]-[NapDNAbp]-[COOH];

[NH2]-[UGI]-[nucleic acid-editing enzyme or domain]-[NapDNAbp]-[second NapDNAbp protein]-[COOH],

[NH2]-[UGI]-[second NapDNAbp protein]-[NapDNAbp]-[nucleic acid-editing enzyme or domain]-[COOH];

[NH2]-[UGI]-[NapDNAbp]-[nucleic acid-editing enzyme or domain]-[second NapDNAbp protein]-[COOH];

[NH2]-[UGI]-[second NapDNAbp protein]-[nucleic acid-editing enzyme or domain]-[NapDNAbpHCOOH];

[NH2]-[nucleic acid-editing enzyme or domain]-[NapDNAbp]-[second NapDNAbp protein]-[UGI]-[COOH];

[NH2]-[second NapDNAbp protein]-[NapDNAbp]-[nucleic acid-editing enzyme or domain]-[UGI]-[COOH];

[NH2]-[NapDNAbp]-[nucleic acid-editing enzyme or domain]-[second NapDNAbp protein]-[UGI]-[COOH]; or

[NH2]-[second NapDNAbp protein]-[nucleic acid-editing enzyme or domain]-[NapDNAbp]-[UGI]-[COOH];

wherein N¾ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, the "]-[" used in the general architecture above indicates the presence of an optional linker sequence. In some embodiments, the second NapDNAbp is a dCas9 protein. In some examples, the general architecture of exemplary Cas9 fusion proteins provided herein comprises a structure as shown in Figure 3. It should be appreciated that any of the proteins provided in any of the general architectures of exemplary Cas9 fusion proteins may be connected by one or more of the linkers provided herein. In some embodiments, the linkers are the same. In some embodiments, the linkers are different. In some embodiments, one or more of the proteins provided in any of the general architectures of exemplary Cas9 fusion proteins are not fused via a linker. In some embodiments, the fusion proteins further comprise a nuclear targeting sequence, for example a nuclear localization sequence. In some embodiments, fusion proteins provided herein further comprise a nuclear localization sequence (NLS). In some embodiments, the NLS is fused to the N-terminus of the fusion protein. In some embodiments, the NLS is fused to the C-terminus of the fusion protein. In some embodiments, the NLS is fused to the N-terminus of the third protein. In some embodiments, the NLS is fused to the C-terminus of the third protein. In some embodiments, the NLS is fused to the N-terminus of the Cas9 protein. In some embodiments, the NLS is fused to the C-terminus of the Cas9 protein. In some embodiments, the NLS is fused to the N-terminus of the nucleic acid-editing enzyme or domain. In some embodiments, the NLS is fused to the C-terminus of the nucleic acid-editing enzyme or domain. In some embodiments, the NLS is fused to the N-terminus of the UGI protein. In some embodiments, the NLS is fused to the C-terminus of the UGI protein. In some embodiments, the NLS is

fused to the fusion protein via one or more linkers. In some embodiments, the NLS is fused to the fusioin protein without a linker

Uracil glycosylase inhibitorfusion proteins

[00346] Some aspects of the disclosure relate to fusion proteins that comprise a uracil glycosylase inhibitor (UGI) domain. In some embodiments, any of the fusion proteins provided herein that comprise a Cas9 domain *(e.g.,* a nuclease active Cas9 domain, a nuclease inactive dCas9 domain, or a Cas9 nickase) may be further fused to a UGI domain either directly or via a linker. Some aspects of this disclosure provide deaminase-dCas9 fusion proteins, deaminase-nuclease active Cas9 fusion proteins and deaminase-Cas9 nickase fusion proteins with increased nucleobase editing efficiency. Without wishing to be bound by any particular theory, cellular DNA-repair response to the presence of U:G heteroduplex DNA may be responsible for the decrease in nucleobase editing efficiency in cells. For example, uracil DNA glycosylase (UDG) catalyzes removal of U from DNA in cells, which may initiate base excision repair, with reversion of the U:G pair to a C:G pair as the most common outcome. As demonstrated in the Examples below, Uracil DNA Glycosylase Inhibitor (UGI) may inhibit human UDG activity. Thus, this disclosure contemplates a fusion protein comprising dCas9-nucleic acid editing domain futher fused to a UGI domain . This disclosure also contemplates a fusion protein comprising a Cas9 nickase-nucleic acid editing domain further fused to a UGI domain. It should be understood that the use of a UGI domain may increase the editing efficiency of a nucleic acid editing domain that is capable of catalyzing a C to U change. For example, fusion proteins comprising a UGI domain may be more efficient in deaminating C residues. In some embodiments, the fusion protein comprises the structure:

[deaminase]-[optional linker sequence]-[dCas9]-[optional linker sequence]-[UGI];
[deaminase]-[optional linker sequence]-[UGI]-[optional linker sequence]-[dCas9];
[UGI]-[optional linker sequence]-[deaminase]-[optional linker sequence]-[dCas9];
[UGI]-[optional linker sequence]-[dCas9]-[optional linker sequence]-[deaminase];
[dCas9]-[optional linker sequence]-[deaminase]-[optional linker sequence]-[UGI];
[dCas9]-[optional linker sequence]-[UGI]-[optional linker sequence]-[deaminase];
[deaminase]-[optional linker sequence]-[dCas9]-[optional linker sequence]-[first UGI]-[optional linker sequence]-[second UGI];
[deaminase]-[optional linker sequence]-[first UGI]-[optional linker seqeunce]-[second UGI]-[optional linker sequence]-[dCas9];
[first UGI]-[optional linker sequence]-[second UGI]-[optional linker sequence]-

[deaminase]-[optional linker sequence]-[dCas9];

[first UGI]-[optional linker sequence]-[second UGi]-[optional linker sequence]-[dCas9]-[optional linker sequence]-[deaminase];

[dCas9]-[optional linker sequence]-[deaminase]-[optional linker sequence]-[first UGI]-[optional linker sequence]-[second UGI]; or

[dCas9]-[optional linker sequence]-[first UGI]-[optional linker sequence]-[second UGI]-[optional linker sequence]-[deaminase].

[00347] In other embodiments, the fusion protein comprises the structure:

[deaminase]-[optional linker sequence]-[Cas9 nickase]-[optional linker sequence]-[UGI];

[deaminase]-[optional linker sequence]-[UGI]-[optional linker sequence]-[Cas9 nickase];

[UGI]-[optional linker sequence]-[deaminase]-[optional linker sequence]-[Cas9 nickase];

[UGI]-[optional linker sequence]-[Cas9 nickase]-[optional linker sequence]-[deaminase];

[Cas9 nickase]-[optional linker sequence]-[deaminase]-[optional linker sequence]-[UGI]; [Cas9 nickase]-[optional linker sequence]-[UGI]-[optional linker sequence]-[deaminase]

[deaminase]-[optional linker sequence]-[Cas9 nickase]-[optional linker sequence]-[first UGI]-[optional linker sequence]-[second UGI];

[deaminase]-[optional linker sequence]-[first UGI]-[optional linker seqeunce]-[second UGI]-[optional linker sequence]-[Cas9 nickase];

[first UGI]-[optional linker sequence]-[second UGI]-[optional linker sequence]-[deaminase]-[optional linker sequence]-[Cas9 nickase];

[first UGI]-[optional linker sequence]-[second UGi]-[optional linker sequence]-[Cas9 nickase]-[optional linker sequence]-[deaminase];

[Cas9 nickase]-[optional linker sequence]-[deaminase]-[optional linker sequence]-[first UGI]-[optional linker sequence]-[second UGI]; or

[Cas9 nickase]-[optional linker sequence]-[first UGI]-[optional linker sequence]-[second UGI]-[optional linker sequence]-[deaminase].

[00348] It should be appreciated that any of the fusion proteins described above may be comprised of (i) a nucleic acid programmable DNA binding protein (napDNAbp); (ii) a cytidine deaminase domain; and (iii) two or more UGI domains, wherein the two or more UGI

domains may be adjacent (e.g., [first UGI]-[second UGI], wherein "-" is an optional linker) to one another in the construct, or the two or more UGI domains may be separated by the napDNAbp of (i) and/or the cytidine deaminase domain of (ii) (e.g., [first UGI]-[deaminase]-[second UGI], [first UGI]-[napDNAbp]-[second UGI], [first UGI]-[deaminase]-[napDNAbp]-[second UGI], ect., wherein "-" is an optional linker).

[00349] In another aspect, the fusion protein comprises: (i) a Cas9 enzyme or domain; (ii) a nucleic acid-editing enzyme or domain (e.g., a second protein) (e.g., a cytidine deaminase domain); (iii) a first uracil glycosylase inhibitor domain (UGI) (e.g., a third protein); and (iv) a second uracil glycosylase inhibitor domain (UGI) (e.g., a fourth protein). The first and second uracil glycosylase inhibitor domains (UGIs) may be the same or different. In some embodiments, the Cas9 domain (e.g., the first protein) and the deaminase (e.g., the second protein) are fused via a linker. In some embodiments, the Cas9 domain is fused to the C-terminus of the deaminase. In some embodiments, the Cas9 protein (e.g., the first protein) and the first UGI domain (e.g., the third protein) are fused via a linker (e.g., a second linker). In some embodiments, the first UGI domain is fused to the C-terminus of the Cas9 protein. In some embodiments, the first UGI domain (e.g., the third protein) and the second UGI domain (e.g., the forth protein) are fused via a linker (e.g., a third linker). In some embodiments, the second UGI domain is fused to the C-terminus of the first UGI domain. In some embodiments, the linker comprises a $(GGGGS)_n$ (SEQ ID NO: 607), a $(G)_n$ (SEQ ID NO: 608), an $(EAAAK)_n$ (SEQ ID NO: 609), a $(GGS)_n$ (SEQ ID NO: 610), $(SGGS)_n$ (SEQ ID NO: 606), a SGSETPGTSESATPES (SEQ ID NO: 604), a $SGGS(GGS)_n$ (SEQ ID NO: 612), a SGGSSGGSSGS ETPGTS ESATPES SGGSSGGS (SEQ ID NO: 605), or an $(XP)_n$ (SEQ ID NO: 611) motif, or a combination of any of these, wherein n is independently an integer between 1 and 30. In some embodiments, the first linker comprises an amino acid sequence of 1-50 amino acids. In some embodiments, the first linker comprises an amino acid sequence of 1-40 amino acids. In some embodiments, the first linker comprises an amino acid sequence of 1-35 amino acids. In some embodiments, the first linker comprises an amino acid sequence of 1-30 amino acids. In some embodiments, the first linker comprises an amino acid sequence of 1-20 amino acids. In some embodiments, the first linker comprises an amino acid sequence of 10-20 amino acids. In some embodiments, the first linker comprises an amino acid sequence of 30-40 amino acids. In some embodiments, the first linker comprises an amino acid sequence of 14, 16, or 18 amino acids. In some embodiments, the first linker comprises an amino acid sequence of 16 amino acids. In some embodiments, the first linker comprises an amino acid sequence of 30, 32, or 34 amino acids. In some embodiments, the first linker

comprises an amino acid sequence of 32 amino acids. In some embodiments, the first linker comprises a SGSETPGTSESATPES (SEQ ID NO: 604) motif. In some embodiments, the first linker comprises a SGGSSGGSSGS ETPGTS E SATPES SGGSSGGS (SEQ ID NO: 605) motif. In some embodiments, the second linker comprises comprises an amino acid sequence of 1-50 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 1-40 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 1-35 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 1-30 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 1-20 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 2-20 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 2-10 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 10-20 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 2, 4, or 6 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 7, 9, or 11 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 14, 16, or 18 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 4 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 9 amino acids. In some embodiments, the second linker comprises an amino acid sequence of 16 amino acids. In some embodiments, the second linker comprises a $(SGGS)_n$ (SEQ ID NO: 606) motif, wherein n is an integer between 1 and 30, inclusive. In some embodiments, the second linker comprises a $(SGGS)_n$ (SEQ ID NO: 606) motif, wherein n is 1. In some embodiments, the second linker comprises a $SGGS(GGS)_n$ (SEQ ID NO: 612) motif, wherein n is an integer between 1 and 30, inclusive. In some embodiments, the second linker comprises a $SGGS(GGS)_n$ (SEQ ID NO: 612) motif, wherein n is 2. In some embodiments, the third linker comprises comprises an amino acid sequence of 1-50 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 1-40 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 1-35 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 1-30 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 1-20 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 2-20 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 2-10 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 10-20 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 2, 4, or 6 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 7, 9,

or 11 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 14, 16, or 18 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 4 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 9 amino acids. In some embodiments, the third linker comprises an amino acid sequence of 16 amino acids. In some embodiments, the third linker comprises a $(SGGS)_n$ (SEQ ID NO: 606) motif, wherein n is an integer between 1 and 30, inclusive. In some embodiments, the third linker comprises a $(SGGS)_n$ (SEQ ID NO: 606) motif, wherein n is 1. In some embodiments, the third linker comprises a $SGGS(GGS)_n$ (SEQ ID NO: 612)motif, wherein n is an integer between 1 and 30, inclusive. In some embodiments, the third linker comprises a $SGGS(GGS)_n$ (SEQ ID NO: 612) motif, wherein n is 2.

[00350] In some embodiments, the fusion protein comprises the structure:

[deaminase]-[optional linker sequence]-[dCas9]-[optional linker sequence]-[first UGI]-[optional linker sequence]-[second UGI];

[deaminase]-[optional linker sequence]-[Cas9 nickase]-[optional linker sequence]-[first UGI]-[optional linker sequence]-[second UGI]; or

[deaminase]-[optional linker sequence]-[Cas9]-[optional linker sequence]-[first UGI]-[optional linker sequence]-[second UGI].

[00351] In another aspect, the fusion protein comprises: (i) a Cas9 enzyme or domain; (ii) a nucleic acid-editing enzyme or domain (e.g., a second protein) (e.g., a cytidine deaminase domain); (iii) more than two uracil glycosylase inhibitor (UGI) domains.

[00352] In some embodiments, the fusion proteins provided herein do not comprise a linker sequence. In some embodiments, one or both of the optional linker sequences are present. In some embodiments, one, two, or three of the optional linker sequences are present.

[00353] In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker sequence. In some embodiments, the fusion proteins comprising a UGI further comprise a nuclear targeting sequence, for example a nuclear localization sequence. In some embodiments, fusion proteins provided herein further comprise a nuclear localization sequence (NLS). In some embodiments, the NLS is fused to the N-terminus of the fusion protein. In some embodiments, the NLS is fused to the C-terminus of the fusion protein. In some embodiments, the NLS is fused to the N-terminus of the UGI protein. In some embodiments, the NLS is fused to the C-terminus of the UGI protein. In some embodiments, the NLS is fused to the N-terminus of the Cas9 protein. In some embodiments, the NLS is fused to the C-terminus of the Cas9 protein. In some embodiments, the NLS is fused to the N-terminus of the deaminase. In some embodiments,

the NLS is fused to the C-terminus of the deaminase. In some embodiments, the NLS is fused to the N-terminus of the second Cas9. In some embodiments, the NLS is fused to the C-terminus of the second Cas9. In some embodiments, the NLS is fused to the fusion protein via one or more linkers. In some embodiments, the NLS is fused to the fusioin protein without a linker. In some embodiments, the NLS comprises an amino acid sequence of any one of the NLS sequences provided or referenced herein. In some embodiments, the NLS comprises an amino acid sequence as set forth in SEQ ID NO: 614or SEQ ID NO: 615.

[00354] In some embodiments, a UGI domain comprises a wild-type UGI or a UGI as set forth in SEQ ID NO: 134. In some embodiments, the UGI proteins provided herein include fragments of UGI and proteins homologous to a UGI or a UGI fragment. For example, in some embodiments, a UGI domain comprises a fragment of the amino acid sequence set forth in SEQ ID NO: 134. In some embodiments, a UGI fragment comprises an amino acid sequence that comprises at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid sequence as set forth in SEQ ID NO: 134. In some embodiments, a UGI comprises an amino acid sequence homologous to the amino acid sequence set forth in SEQ ID NO: 134 or an amino acid sequence homologous to a fragment of the amino acid sequence set forth in SEQ ID NO: 134. In some embodiments, proteins comprising UGI or fragments of UGI or homologs of UGI or UGI fragments are referred to as "UGI variants." A UGI variant shares homology to UGI, or a fragment thereof. For example a UGI variant is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, at least 99.5% identical, or at least 99.9% identical to a wild type UGI or a UGI as set forth in SEQ ID NO: 134. In some embodiments, the UGI variant comprises a fragment of UGI, such that the fragment is at least 70% identical, at least 80% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, at least 99.5% identical, or at least 99.9% to the corresponding fragment of wild-type UGI or a UGI as set forth in SEQ ID NO: 134. In some embodiments, the UGI comprises the following amino acid sequence:

>splP14739IUNGI_BPPB2 Uracil-DNA glycosylase inhibitor
MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS
DAPEYKPWALVIQDSNGENKIKML (SEQ ID NO: 134)

[00355] Suitable UGI protein and nucleotide sequences are provided herein and additional suitable UGI sequences are known to those in the art, and include, for example, those published in Wang et al., Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. J. Biol. Chem. 264: 1163-1171(1989); Lundquist et al., Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein. Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase. J. Biol. Chem. 272:21408-21419(1997); Ravishankar et al., X-ray analysis of a complex of Escherichia coli uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. Nucleic Acids Res. 26:4880-4887(1998); and Putnam et al., Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. J. Mol. Biol. 287:331-346(1999), the entire contents of each are incorporated herein by reference.

[00356] It should be appreciated that additional proteins may be uracil glycosylase inhibitors. For example, other proteins that are capable of inhibiting *(e.g.,* sterically blocking) a uracil-DNA glycosylase base-excision repair enzyme are within the scope of this disclosure. Additionally, any proteins that block or inhibit base-excision repair as also within the scope of this disclosure. In some embodiments, the fusion proteins described herein comprise one UGI domain. In some emobidments, the fusion proteins described herein comprise more than one UGI domain. In some embodiments, the fusion proteins described herein comprise two UGI domains. In some embodiments, the fusion proteins described herein comprise more than two UGI domains. In some embodiments, a protein that binds DNA is used. In another embodiment, a substitute for UGI is used. In some embodiments, a uracil glycosylase inhibitor is a protein that binds single-stranded DNA. For example, a uracil glycosylase inhibitor may be a Erwinia tasmaniensis single-stranded binding protein. In some embodiments, the single-stranded binding protein comprises the amino acid sequence (SEQ ID NO: 135). In some embodiments, a uracil glycosylase inhibitor is a protein that binds uracil. In some embodiments, a uracil glycosylase inhibitor is a protein that binds uracil in DNA. In some embodiments, a uracil glycosylase inhibitor is a catalytically inactive uracil DNA-glycosylase protein. In some embodiments, a uracil glycosylase inhibitor is a catalytically inactive uracil DNA-glycosylase protein that does not excise uracil from the DNA. For example, a uracil glycosylase inhibitor is a UdgX. In some embodiments, the UdgX comprises the amino acid sequence (SEQ ID NO: 136). As another example, a uracil glycosylase inhibitor is a catalytically inactive UDG. In some embodiments, a catalytically

inactive UDG comprises the amino acid sequence (SEQ ID NO: 137). It should be appreciated that other uracil glycosylase inhibitors would be apparent to the skilled artisan and are within the scope of this disclosure. In some embodiments, a uracil glycosylase inhibitor is a protein that is homologous to any one of SEQ ID NOs: 135-137 or 143-148. In some embodiments, a uracil glycosylase inhibitor is a protein that is at least 50% identical, at least 55% identical at least 60% identical, at least 65% identical, at least 70% identical, at least 75% identical, at least 80% identical at least 85% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 98% identical, at least 99% identical, or at least 99.5% identical to any one of SEQ ID NOs: 135-137 or 143-148.

Erwinia tasmaniensis SSB (themostable single-stranded DNA binding protein)
MASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKQTGETKEKTEWH
RVVLFGKLAEVAGEYLRKGSQVYIEGALQTRKWTDQAGVEKYTTEVVVNVGGTMQ
MLGGRSQGGGASAGGQNGGSNNGWGQPQQPQGGNQFSGGAQQQARPQQQPQQNN
APANNEPPIDFDDDIP (SEQ ID NO: 135)

UdgX (binds to Uracil in DNA but does not excise)
MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGEQPG
DKEDLAGLPFVGPAGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKTP
SRTEVVACRPWLIAEMTSVEPDVVVLLGATAAKALLGNDFRVTQHRGEVLHVDDVP
GDPALVATVHPSSLLRGPKEERES AFAGLVDDLRV AADVRP (SEQ ID NO: 136)

UDG (catalytically inactive human UDG, binds to Uracil in DNA but does not excise)
MIGQKTLYS FFSPSPARKRH APSPEPAVQGTG VAGVPEES GDAAAIP AKKAP AGQEEP
GTPPSSPLSAEQLDRIQRNKAAALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKLMG
FVAEERKHYTVYPPPHQVFTWTQMCDIKDVKVVILGQEPYHGPNQAHGLCFSVQRPV
PPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANSHKERG
WEQFTD AVVSWLNQNS NGLVFLLWGS YAQKKGS AIDRKRHH VLQT AHPSPLS VYRG
FFGCRHFS KTNELLQKS GKKPID WKEL (SEQ ID NO: 137)

[00357] Additional single-stranded DNA binding proteins that can be used as a UGI are shown below. It should be appreciated that other single-stranded binding proteins may be used as a UGI, for example those described in Dickey TH, Altschuler SE, Wuttke DS. Single-

stranded DNA-binding proteins :multiple domains for multiple functions. Structure. 2013 Jul 2;21(7): 1074-84.

doi: 10.1016/j.str.2013.05.013. Review.; Marceau AH. Functions of single-strand DNA-binding proteins in DNA replication, recombination, and repair. Methods Mol Biol. 2012;922:1-21. doi:

10. 1007/978- l-62703-032-8_l.; Mijakovic, Ivan, *et al* ; Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine. Nucleic Acids Res 2006; 34 (5): 1588-1596. doi: 10.1093/nar/gkj514; Mumtsidu E, Makhov AM, Konarev PV, Svergun DI, Griffith JD, Tucker PA. Structural features of the single-stranded DNA-binding protein of Epstein-Barrvirus. J Struct Biol. 2008 Feb;161(2):172-87. Epub 2007 Nov 1; Nowak M, Olszewski M, Śpibida M, Kur J. Characterization of single-strandedDNA-binding proteins from the psychrophilic bacteria Desulfotalea psychrophila, Flavobacterium psychrophilum, Psychrobacter arcticus, Psychrobactercryohalolentis, Psychromonas ingrahamii, Psychroflexus torquis, and

Photobacterium profundum. BMC Microbiol. 2014 Apr 14;14:91. doi: 10.1186/1471-2180-14-91; Tone T, Takeuchi A, Makino O. Single-stranded DNA binding protein Gp5 of Bacillus subtilis phage Φ29 is required for viral DNA replication in growth-temperature dependent fashion. Biosci Biotechnol Biochem. 2012;76(12):2351-3. Epub 2012 Dec 7; Wold. REPLICATION PROTEIN A:A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism. Annual Review of Biochem. 1997; 66:61-92. doi: 10.1146/annurev.biochem.66.1.61; Wu Y, Lu J, Kang T. Human single-stranded DNA binding proteins: guardians of genome stability. Acta Biochim Biophys Sin (Shanghai). 2016 Jul;48(7):671-7. doi: 10.1093/abbs/gmw044. Epub 2016 May 23. Review; the entire contents of each are hereby incorporated by reference.

mtSSB - SSBP1 single stranded DNA binding protein 1 [Homo sapiens (human)] (UniProtKB: Q04837; NP_001243439.1)

MFRRPVLQVLRQFVRHESETTTSLVLERSLNRVHLLGRVGQDPVLRQVEGKNPVTIFS
LATNEMWRSGDSEVYQLGDVSQKTTWHRISVFRPGLRDVAYQYVKKGSRIYLEGKI
DYGEYMDKNNVRRQATTIIADNIIFLSDQTKEKE (SEQ ID NO: 138)

Single-stranded DNA-binding protein 3 isoform A [Mus musculus] (UniProtKB - Q9D032-1; NCBI Ref: NP_076161.2)

MFAKGKGSAVPSDGQAREKLALYVYEYLLHVGAQKSAQTFLSEIRWEKNITLGEPPG
FLHSWWCVFWDLYCAAPERRDTCEHSSEAKAFHDYSAAAAPSPVLGNIPPNDGMPG
GPIPPGFFQGPPGSQPSPHAQPPPHNPSSMMGPHSQPFMSPRYAGGPRPPIRMGNQPPG
GVPGTQPLLPNSMDPTRQQGHPNMGGSMQRMNPPRGMGPMGPGPQNYGSGMRPPP
NSLGPAMPGINMGPG AGRPWPNPNS ANSIPYSSSSPGTYVGPPGGGGPPGTPIMPS PAD

STNSSDNIYTMINPVPPGGSRSNFPMGPGSDGPMGGMGGMEPHHMNGSLGSGDIDGL
PKNSPNNISGISNPPGTPRDDGELGGNFLHS FQNDNYSPSMTMSV (SEQ ID NO: 139)

RPA 1 - Replication protein A 70 kDa DNA-binding subunit (UniProtKB: P27694; NCBI Ref: NM_002945.3)

MVGQLSEGAIAAIMQKGDTNIKPILQVINIRPITTGNSPPRYRLLMSDGLNTLSSFMLAT
QLNPLVEEEQLS SNCVCQIHRFIVNTLKD GRRVVILMELE VLKSAEAVGVKIGNP VPY
NE
GLGQPQVAPP APAASPAASSRPQPQNGS SGMGSTVSKAYGASKTFGKA AGPSLSHTS
GGTQSKVVPIASLTPYQSKWTIC ARVTNKS QIRTWSNSRGEGKLFS LELVDES GEIRAT
AFNEQVDKFFPLIEVNKVYYFSKGTLKIANKQFTAVKNDYEMTFNNETSVMPCEDDH
HLPTVQFDFTGIDDLENKSKDSLVDIIGICKSYEDATKITVRSNNREVAKRNIYLMDTS
GKVVTATLWGEDADKFDGSRQPVLAIKGARVSDFGGRSLSVLSSSTIIANPDIPEAYKL
RGWFDAEGQALDGVSISDLKSGGVGGSNTNWKTLYEVKSENLGQGDKPDYFSSVAT
VVYLRKENCMYQACPTQDCNKKVIDQQNGLYRCEKCDTEFPNFKYRMILSVNIADFQ
ENQWVTCFQESAEAILGQNAAYLGELKDKNEQAFEEVFQNANFRSFIFRVRVKVETY
NDESRIKATVMDVKPVDYREYGRRLVMS IRRSALM (SEQ ID NO: 140)

RPA 2 - Replication protein A 32 kDa subunit (UniProtKB: P15927; NCBI Ref: NM_002946)

MWNSGFESYGSSSYGGAGGYTQSPGGFGS PAPSQAEKKSRARAQHIVPCTIS QLLSAT
LVDEVFRIGNVEISQVTIVGIIRHAEKAPTNIVYKIDDMTAAPMDVRQWVDTDDTSSE
NTVVPPETYVKVAGHLRSFQNKKSLVAFKIMPLEDMNEFTTHILEVINAHMVLSKAN
SQPSAGRAPISNPGMSEAGNFGGNSFMPANGLTVAQNQVLNLIKACPRPEGLNFQDL
KNQLKHMSVSSIKQAVDFLSNEGHIYSTVDDDHFKSTDAE (SEQ ID NO: 141)

RPA 3 - Replication protein A 14 kDa subunit (UniProtKB: P35244; NCBI Ref: NM_002947.4)

MVDMMDLPRSRINAGMLAQFIDKPVCFVGRLEKIHPTGKMFILSDGEGKNGTIELMEP
LDEEISGIVEVVGRVTAKATILCTSYVQFKEDSHPFDLGLYNEAVKIIHDFPQFYPLGIV
QH
D (SEQ ID NO: 142)

Bacterial single-stranded DNA-binding proteins:

ssbA - single-stranded DNA-binding protein [Bacillus subtilis subsp. subtilis str. 168] (UniProtKB: P37455; NCBI Ref:)

MLNRVVLVGR LTKDPELRYT PNGAAVATFT LAVNRTFTNQ SGEREADFIN
CVTWRRQAEN VANFLKKGSL AGVDGRLQTR NYENQQGQRV FVTEVQAESV
QFLEPKNGGG SGSGGYNEGN SGGGQYFGGG QNDNPFGGNQ NNQRRNQGNS
FNDDPFANDG KPIDISDDDLPF (SEQ ID NO: 143)

Single-stranded DNA-binding protein 2 [Streptomyces coelicolor A3(2)] (UniProtKB: Q9X8U3; NCBI Ref: NP_628093.1)

MAGETVITVVGNLVDDPELRFTPSGAAVAKFRVASTPRTFDRQTNEWKDGESLFLTC
SVWRQAAENVAESLQRGMRVIVQGRLKQRSYEDREGVKRTVYELDVDEVGASLRSA
TAKVTKTSGQGRGGQGGYGGGGGGQGGGGWGGGPGGGQQGGGAPADDPWATGG
APAGGQQGGGGQGGGGWGGGSGGGGGYSDEPPF (SEQ ID NO: 144)

Single-stranded DNA-binding protein [Streptococcus pneumoniae R6] (UniProtKB: P66855; NCBI Ref: NP_358988.1)

MINNVVLVGRMTRDAELRYTPSNVAVATFTLAVNRTFKSQNGEREADFINVVMWRQ
QAENLANWAKKGSLIGVTGRIQTRSYDNQQGQRVYVTEVVAENFQMLESRSVREGH
TGGAYSAPTANYSAPTNSVPDFSRNENPFGATNPLDISDDDLPF (SEQ ID NO: 145)

Viral single-stranded DNA-binding proteins:

Single-stranded DNA-binding protein [Human alphaherpesvirus 1] (UniProtKB: P04296; NCBI Ref: YP_009137104.1)

METKPKTATTIKVPPGPLGYVYARACPSEGIELLALLSARSGDSDVAVAPLVVGLTVE
SGFEANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRL
CERARRHFGFSDYTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAVCINN
TFLHLGGSDKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEPFNANHRSIGENFTYPLPFFN
RPLNRLLFEAVVGPAAVALRCRNVDAVARAAAHLAFDENHEGAALPADITFTAFEAS
QGKTPRGGRDGGGKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPTDISAWPL
FEGQDTAAARANAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSKPS
FYRFFLVPGTHVAANPQVDREGHVVPGFEGRPTAPLVGGTQEFAGEHLAMLCGFSPA
LLAKMLFYLERCDGGVIVGRQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTTLM
RLRARHPKFASAARGAIGVFGTMNSMYSDCDVLGNYAAFSALKRADGSETARTIMQ
ETYRAATERVMAELETLQYVDQAVPTAMGRLETIITNREALHTVVNNVRQVVDREV
EQLMRNLVEGRNFKFRDGLGEANHAMSLTLDPYACGPCPLLQLLGRRSNLAVYQDL
ALSQCHGVFAGQSVEGRNFRNQFQPVLRRRVMDMFNNGFLSAKTLTVALSEGAAIC
APSLTAGQTAPAESSFEGDVARVTLGFPKELRVKSRVLFAGASANASEAAKARVASL
QSAYQKPDKRVDILLGPLGFLLKQFHAAIFPNGKPPGSNQPNPQWFWTALQRNQLPA
RLLSREDIETIAFIKKFSLDYGAINFINLAPNNVSELAMYYMANQILRYCDHSTYFINTL
TAIIAGSRRPPSVQAAAAAWSAQGGAGLEAGARALMDAVDAHPGAWTSMFASCNLLR
PVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLMGGKNACPLLIFDRTRKFVL
ACPRAGFVCAASSLGGGAHESSLCEQLRGIISEGGAAVASSVFVATVKSLGPRTQQLQI
EDWLALLEDEYLSEEMMELTARALERGNGEWSTDAALEVAHEAEALVSQLGNAGE
VFNFGDFGCEDDNATPFGGPGAPGPAFAGRKRAFHGDDPFGEGPPDKKGDLTLDML
(SEQ ID NO: 146)

Single-stranded DNA-binding protein from Bacillus virus phi29 (UniProtKB: Q38504.1; NCBI Ref: YP_002004532.1)

MENTNIVKATFDTETLEGQIKIFNAQTGGGQSFKNLPDGTIIEANAIAQYKQVSDTYGD
AKEETVTTIFAADGSLYSAISKTVAEAASDLIDLVTRHKLETFKVKVVQGTSSKGNVF
FSLQLSL (SEQ ID NO: 147)

Single stranded DNA binding protein [Burkholderia virus DC1] (UniProtKB: I6NRL7; NCBI Ref: YP_006589943.1)

MASVNKVILVGNLGADPETRYLPSGDAISNIRLATTDRYKDKASGEMKESTEWHRVS FFGRLAEIVDEYLRKGAPVYIEGRIRTRKWQDNAGQDRYTTEIVAEKMQMLGDRRDG GERQQRAPQQQQQRTQRNGYADATGRAQPSQRPAAGGGFDEMDDDIPF (SEQ ID NO: 148)

[00358] In some embodiments, the nucleic acid editing domain is a deaminase domain. In some embodiments, the deaminase is a cytosine deaminase or a cytidine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBEC 1 deaminase. In some embodiments, the deaminase is an APOBEC2 deaminase. In some embodiments, the deaminase is an APOBEC3 deaminase. In some embodiments, the deaminase is an APOBEC3A deaminase. In some embodiments, the deaminase is an APOBEC3B deaminase. In some embodiments, the deaminase is an APOBEC3C deaminase. In some embodiments, the deaminase is an APOBEC3D deaminase. In some embodiments, the deaminase is an APOBEC3E deaminase. In some embodiments, the deaminase is an APOBEC3F deaminase. In some embodiments, the deaminase is an APOBEC3G deaminase. In some embodiments, the deaminase is an APOBEC3H deaminase. In some embodiments, the deaminase is an APOBEC4 deaminase. In some embodiments, the deaminase is an activation-induced deaminase (AID). In some embodiments, the demianse is a rat APOBEC1 (SEQ ID NO: 74). In some embodiments, the deminase is a human APOBEC 1 (SEQ ID No: 76). In some embodiments, the deaminase is a *Petromyzon marinus* cytidine deaminase 1 (pmCDAl). In some embodiments, the deminase is a human APOBEC3G (SEQ ID NO: 60). In some embodiments, the deaminase is a fragment of the human APOBEC3G (SEQ ID NO: 83). In some embodiments, the deaminase is a human APOBEC3G variant comprising a D316R_D317R mutation (SEQ ID NO: 82). In some embodiments, the deaminase is a frantment of the human APOBEC3G and comprising mutations corresponding to the D316R_D317R mutations in SEQ ID NO: 60 (SEQ ID NO: 84).

[00359] In some embodiments, the linker comprises a $(GGGS)_n$ (SEQ ID NO: 613), $(GGGGS)_n$ (SEQ ID NO: 607), a $(G)_n$ (SEQ ID NO: 608), an $(EAAAK)_n$ (SEQ ID NO: 609), a $(GGS)_n$ (SEQ ID NO: 610), an SGSETPGTSESATPES (SEQ ID NO: 604), or an $(XP)_n$ (SEQ ID NO: 611) motif, or a combination of any of these, wherein n is independently an integer between 1 and 30.

[00360] Suitable UGI protein and nucleotide sequences are provided herein and additional suitable UGI sequences are known to those in the art, and include, for example, those published in Wang *et al.,* Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. *J. Biol. Chem.* 264: 1163-1171(1989); Lundquist *et al.,* Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein. Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase. *J. Biol. Chem.* 272:21408-21419(1997); Ravishankar *et al.,* X-ray analysis of a complex of *Escherichia coli* uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. *Nucleic Acids Res.* 26:4880-4887(1998); and Putnam *et al,* Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. *J. Mol. Biol.* 287:331-346(1999), the entire contents of which are incorporated herein by reference. In some embodiments, the optional linker comprises a $(GGS)_n$ (SEQ ID NO: 610) motif, wherein n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 19, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20. In some embodiments, the optional linker comprises a $(GGS)_n$ (SEQ ID NO: 610) motif, wherein n is 1, 3, or 7. In some embodiments, the optional linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 604), which is also referred to as the XTEN linker in the Examples.

[00361] In some embodiments, a Cas9 nickase may further facilitate the removal of a base on the non-edited strand in an organism whose genome is edited *in vivo.* The Cas9 nickase, as described herein, may comprise a D10A mutation in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. In some embodiments, the Cas9 nickase of this disclosure may comprise a histidine at mutation 840 of SEQ ID NO: 6, or a corresponding residue in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein. Such fusion proteins comprising the Cas9 nickase, can cleave a single strand of the target DNA sequence, *e.g.,* the strand that is not being edited. Without wishing to be bound by any particular theory, this cleavage may inhibit mis-match repair mechanisms that reverse a C to U edit made by the deaminase.

Cas9 complexes with guide RNAs

[00362] Some aspects of this disclosure provide complexes comprising any of the fusion proteins provided herein, and a guide RNA bound to a Cas9 domain *(e.g.,* a dCas9, a nuclease active Cas9, or a Cas9 nickase) of fusion protein.

[00363] In some embodiments, the guide RNA is from 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the guide RNA is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides long. In some embodiments, the guide RNA comprises a sequence of 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the target sequence is a DNA sequence. In some embodiments, the target sequence is a sequence in the genome of a mammal. In some embodiments, the target sequence is a sequence in the genome of a human. In some embodiments, the 3' end of the target sequence is immediately adjacent to a canonical PAM sequence (NGG). In some embodiments, the guide RNA is complementary to a sequence associated with a disease or disorder.

Methods of using Cas9 fusion proteins

[00364] Some aspects of this disclosure provide methods of using the Cas9 proteins, fusion proteins, or complexes provided herein. For example, some aspects of this disclosure provide methods comprising contacting a DNA molecule (a) with any of the the Cas9 proteins or fusion proteins provided herein, and with at least one guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence; or (b) with a Cas9 protein, a Cas9 fusion protein, or a Cas9 protein or fusion protein complex with at least one gRNA as provided herein. In some embodiments, the 3' end of the target sequence is not immediately adjacent to a canonical PAM sequence (NGG). In some embodiments, the 3' end of the target sequence is immediately adjacent to an AGC, GAG, TTT, GTG, or CAA sequence.

[00365] In some embodiments, the target DNA sequence comprises a sequence associated with a disease or disorder. In some embodiments, the target DNA sequence comprises a point mutation associated with a disease or disorder. In some embodiments, the activity of the Cas9 protein, the Cas9 fusion protein, or the complex results in a correction of the point mutation. In some embodiments, the target DNA sequence comprises a T→C point mutation associated with a disease or disorder, and wherein the deamination of the mutant C base results in a sequence that is not associated with a disease or disorder. In some embodiments, the target DNA sequence encodes a protein and wherein the point mutation is in a codon and results in a change in the amino acid encoded by the mutant codon as compared to the wild-type codon. In some embodiments, the deamination of the mutant C results in a change of the amino acid

encoded by the mutant codon. In some embodiments, the deamination of the mutant C results in the codon encoding the wild-type amino acid. In some embodiments, the contacting is *in vivo* in a subject. In some embodiments, the subject has or has been diagnosed with a disease or disorder. In some embodiments, the disease or disorder is cystic fibrosis, phenylketonuria, epidermolytic hyperkeratosis (EHK), Charcot-Marie-Toot disease type 4J, neuroblastoma (NB), von Willebrand disease (vWD), myotonia congenital, hereditary renal amyloidosis, dilated cardiomyopathy (DCM), hereditary lymphedema, familial Alzheimer's disease, HIV, Prion disease, chronic infantile neurologic cutaneous articular syndrome (CINCA), desmin-related myopathy (DRM), a neoplastic disease associated with a mutant PI3KCA protein, a mutant CTNNB 1 protein, a mutant HRAS protein, or a mutant p53 protein.

[00366] Some embodiments provide methods for using the Cas9 DNA editing fusion proteins provided herein. In some embodiments, the fusion protein is used to introduce a point mutation into a nucleic acid by deaminating a target nucleobase, *e.g.*, a C residue. In some embodiments, the deamination of the target nucleobase results in the correction of a genetic defect, *e.g.*, in the correction of a point mutation that leads to a loss of function in a gene product. In some embodiments, the genetic defect is associated with a disease or disorder, *e.g.*, a lysosomal storage disorder or a metabolic disease, such as, for example, type I diabetes. In some embodiments, the methods provided herein are used to introduce a deactivating point mutation into a gene or allele that encodes a gene product that is associated with a disease or disorder. For example, in some embodiments, methods are provided herein that employ a Cas9 DNA editing fusion protein to introduce a deactivating point mutation into an oncogene *(e.g.*, in the treatment of a proliferative disease). A deactivating mutation may, in some embodiments, generate a premature stop codon in a coding sequence, which results in the expression of a truncated gene product, *e.g.*, a truncated protein lacking the function of the full-length protein.

[00367] In some embodiments, the purpose of the methods provide herein is to restore the function of a dysfunctional gene via genome editing. The Cas9 deaminase fusion proteins provided herein can be validated for gene editing-based human therapeutics *in vitro, e.g.*, by correcting a disease-associated mutation in human cell culture. It will be understood by the skilled artisan that the fusion proteins provided herein, *e.g.*, the fusion proteins comprising a Cas9 domain and a nucleic acid deaminase domain can be used to correct any single point T -> C or A -> G mutation. In the first case, deamination of the mutant C back to U corrects the mutation, and in the latter case, deamination of the C that is base-paired with the mutant G, followed by a round of replication, corrects the mutation.

[00368] An exemplary disease-relevant mutation that can be corrected by the provided fusion proteins *in vitro* or *in vivo* is the H1047R (A3140G) polymorphism in the PI3KCA protein. The phosphoinositide-3-kinase, catalytic alpha subunit (PI3KCA) protein acts to phosphorylate the 3-OH group of the inositol ring of phosphatidylinositol. The PI3KCA gene has been found to be mutated in many different carcinomas, and thus it is considered to be a potent oncogene.[37] In fact, the A3140G mutation is present in several NCI-60 cancer cell lines, such as, for example, the HCT116, SKOV3, and T47D cell lines, which are readily available from the American Type Culture Collection (ATCC).[38]

[00369] In some embodiments, a cell carrying a mutation to be corrected, *e.g., a* cell carrying a point mutation, *e.g.,* an A3140G point mutation in exon 20 of the PI3KCA gene, resulting in a H1047R substitution in the PI3KCA protein, is contacted with an expression construct encoding a Cas9 deaminase fusion protein and an appropriately designed sgRNA targeting the fusion protein to the respective mutation site in the encoding PI3KCA gene. Control experiments can be performed where the sgRNAs are designed to target the fusion enzymes to non-C residues that are within the PI3KCA gene. Genomic DNA of the treated cells can be extracted, and the relevant sequence of the PI3KCA genes PCR amplified and sequenced to assess the activities of the fusion proteins in human cell culture.

[00370] It will be understood that the example of correcting point mutations in PI3KCA is provided for illustration purposes and is not meant to limit the instant disclosure. The skilled artisan will understand that the instantly disclosed DNA-editing fusion proteins can be used to correct other point mutations and mutations associated with other cancers and with diseases other than cancer including other proliferative diseases.

[00371] The successful correction of point mutations in disease-associated genes and alleles opens up new strategies for gene correction with applications in therapeutics and basic research. Site-specific single-base modification systems like the disclosed fusions of Cas9 and deaminase enzymes or domains also have applications in "reverse" gene therapy, where certain gene functions are purposely suppressed or abolished. In these cases, site-specifically mutating Trp (TGG), Gin (CAA and CAG), or Arg (CGA) residues to premature stop codons (TAA, TAG, TGA) can be used to abolish protein function *in vitro, ex vivo,* or *in vivo.*

[00372] The instant disclosure provides methods for the treatment of a subject diagnosed with a disease associated with or caused by a point mutation that can be corrected by a Cas9 DNA editing fusion protein provided herein. For example, in some embodiments, a method is provided that comprises administering to a subject having such a disease, *e.g.,* a cancer associated with a PI3KCA point mutation as described above, an effective amount of a Cas9

deaminase fusion protein that corrects the point mutation or introduces a deactivating mutation into the disease-associated gene. In some embodiments, the disease is a proliferative disease. In some embodiments, the disease is a genetic disease. In some embodiments, the disease is a neoplastic disease. In some embodiments, the disease is a metabolic disease. In some embodiments, the disease is a lysosomal storage disease. Other diseases that can be treated by correcting a point mutation or introducing a deactivating mutation into a disease-associated gene will be known to those of skill in the art, and the disclosure is not limited in this respect.

[00373] The instant disclosure provides methods for the treatment of additional diseases or disorders, *e.g.*, diseases or disorders that are associated or caused by a point mutation that can be corrected by deaminase-mediated gene editing. Some such diseases are described herein, and additional suitable diseases that can be treated with the strategies and fusion proteins provided herein will be apparent to those of skill in the art based on the instant disclosure. Exemplary suitable diseases and disorders are listed below. It will be understood that the numbering of the specific positions or residues in the respective sequences depends on the particular protein and numbering scheme used. Numbering might be different, *e.g.*, in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species may affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, *e.g.*, by sequence alignment and determination of homologous residues. Exemplary suitable diseases and disorders include, without limitation, cystic fibrosis (see, *e.g.*, Schwank *et al.*, Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell stem cell.* **2013;** 13: 653-658; and Wu et. al., Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell stem cell.* **2013;** 13: 659-662, neither of which uses a deaminase fusion protein to correct the genetic defect); phenylketonuria - *e.g.*, phenylalanine to serine mutation at position 835 (mouse) or 240 (human) or a homologous residue in phenylalanine hydroxylase gene (T>C mutation) - see, *e.g.*, McDonald *et al, Genomics.* **1997;** 39:402-405; Bernard-Soulier syndrome (BSS) - *e.g.*, phenylalanine to serine mutation at position 55 or a homologous residue, or cysteine to arginine at residue 24 or a homologous residue in the platelet membrane glycoprotein IX (T>C mutation) - see, *e.g.*, Noris *et al., British Journal of Haematology.* **1997;** 97: 312-320, and Ali *et al., Hematol.* **2014;** 93: 381-384; epidermolytic hyperkeratosis (EHK) - *e.g.*, leucine to proline mutation at position 160 or 161 (if counting the initiator methionine) or a homologous residue in keratin 1 (T>C mutation) - see, *e.g.*, Chipev *et al., Cell.* **1992;** 70: 821-828, see also accession number P04264 in the UNIPROT database at www[dot]uniprot[dot]org; chronic

obstructive pulmonary disease (COPD) - *e.g.,* leucine to proline mutation at position 54 or 55 (if counting the initiator methionine) or a homologous residue in the processed form of ai-antitrypsin or residue 78 in the unprocessed form or a homologous residue (T>C mutation) - see, *e.g.,* Poller *et al., Genomics.* **1993;** 17: 740-743, see also accession number POlOll in the UNIPROT database; Charcot-Marie-Toot disease type 4J - *e.g.,* isoleucine to threonine mutation at position 41 or a homologous residue in FIG4 (T>C mutation) - see, *e.g.,* Lenk *et al.,* PLoS Genetics. 2011; 7: el002104; neuroblastoma (NB) - *e.g.,* leucine to proline mutation at position 197 or a homologous residue in Caspase-9 (T>C mutation) - see, *e.g.,* Kundu *et al., 3 Biotech.* **2013, 3:225-234;** von Willebrand disease (vWD) - *e.g.,* cysteine to arginine mutation at position 509 or a homologous residue in the processed form of von Willebrand factor, or at position 1272 or a homologous residue in the unprocessed form of von Willebrand factor (T>C mutation) - see, *e.g.,* Lavergne *et al., Br. J. Haematol.* **1992,** see also accession number P04275 in the UNIPROT database; 82: 66-72; myotonia congenital - *e.g.,* cysteine to arginine mutation at position 277 or a homologous residue in the muscle chloride channel gene CLCN1 (T>C mutation) - see, *e.g.,* Weinberger *et al, The J. of Physiology.* **2012;** 590: 3449-3464; hereditary renal amyloidosis - *e.g.,* stop codon to arginine mutation at position 78 or a homologous residue in the processed form of apolipoprotein All or at position 101 or a homologous residue in the unprocessed form (T>C mutation) - see, *e.g.,* Yazaki *et al., Kidney Int.* **2003;** 64: 11-16; dilated cardiomyopathy (DCM) - *e.g.,* tryptophan to Arginine mutation at position 148 or a homologous residue in the FOXD4 gene (T>C mutation), see, *e.g.,* Minoretti et. al., *Int. J. ofMol. Med.* **2007;** 19: 369-372; hereditary lymphedema - *e.g.,* histidine to arginine mutation at position 1035 or a homologous residue in VEGFR3 tyrosine kinase (A>G mutation), see, *e.g.,* Irrthum *et al., Am. J. Hum. Genet.* **2000;** 67: 295-301; familial Alzheimer's disease - *e.g.,* isoleucine to valine mutation at position 143 or a homologous residue in presenilinl (A>G mutation), see, *e.g.,* Gallo et. al., *J. Alzheimer' s disease.* **2011;** 25: 425-431; Prion disease - *e.g.,* methionine to valine mutation at position 129 or a homologous residue in prion protein (A>G mutation) - see, *e.g.,* Lewis et. al., *J. of General Virology.* **2006;** 87: 2443-2449; chronic infantile neurologic cutaneous articular syndrome (CINCA) - *e.g.,* Tyrosine to Cysteine mutation at position 570 or a homologous residue in cryopyrin (A>G mutation) - see, *e.g.,* Fujisawa et. al. *Blood.* **2007;** 109: 2903-2911; and desmin-related myopathy (DRM) - *e.g.,* arginine to glycine mutation at position 120 or a homologous residue in αβ crystallin (A>G mutation) - see, *e.g.,* Kumar *et al., J. Biol. Chem.* **1999;** 274: 24137-24141. The entire contents of all references and database entries is incorporated herein by reference.

[00374] It will be apparent to those of skill in the art that in order to target a Cas9:nucleic acid editing enzyme/domain fusion protein as disclosed herein to a target site, *e.g.*, a site comprising a point mutation to be edited, it is typically necessary to co-express the Cas9:nucleic acid editing enzyme/domain fusion protein together with a guide RNA, *e.g.*, an sgRNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for Cas9 binding, and a guide sequence, which confers sequence specificity to the Cas9:nucleic acid editing enzyme/domain fusion protein. In some embodiments, the guide RNA comprises a structure 5'-[guide sequence]-guuuuagagcuagaaauagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuu u-3' (SEQ ID NO: 618), wherein the guide sequence comprises a sequence that is complementary to the target sequence. The guide sequence is typically 20 nucleotides long. The sequences of suitable guide RNAs for targeting Cas9:nucleic acid editing enzyme/domain fusion proteins to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting Cas9:nucleic acid editing enzyme/domain fusion proteins to specific target sequences are provided below.

Base Editor Efficiency

[00375] Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of modifying a specific nucleotide base without generating a significant proportion of indels. An "indel", as used herein, refers to the insertion or deletion of a nucleotide base within a nucleic acid. Such insertions or deletions can lead to frame shift mutations within a coding region of a gene. In some embodiments, it is desirable to generate base editors that efficiently modify *(e.g.* mutate or deaminate) a specific nucleotide within a nucleic acid, without generating a large number of insertions or deletions *(i.e.*, indels) in the nucleic acid. In certain embodiments, any of the base editors provided herein are capable of generating a greater proportion of intended modifications *(e.g.*, point mutations or deaminations) versus indels. In some embodiments, the base editors provided herein are capable of generating a ratio of intended point mutations to indels that is greater than 1:1. In some embodiments, the base editors provided herein are capable of generating a ratio of intended point mutations to indels that is at least 1.5:1, at least 2:1, at least 2.5:1, at least 3:1, at least 3.5:1, at least 4:1, at least 4.5:1, at least 5:1, at least 5.5:1, at least 6:1, at least 6.5:1, at

least 7:1, at least 7.5:1, at least 8:1, at least 10:1, at least 12:1, at least 15:1, at least 20:1, at least 25:1, at least 30:1, at least 40:1, at least 50:1, at least 100:1, at least 200:1, at least 300:1, at least 400:1, at least 500:1, at least 600:1, at least 700:1, at least 800:1, at least 900:1, or at least 1000:1, or more. The number of intended mutations and indels may be determined using any suitable method, for example the methods used in the below Examples.

[00376] In some embodiments, the base editors provided herein are capable of limiting formation of indels in a region of a nucleic acid. In some embodiments, the region is at a nucleotide targeted by a base editor or a region within 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nucleotide targeted by a base editor. In some embodiments, any of the base editors provided herein are capable of limiting the formation of indels at a region of a nucleic acid to less than 1%, less than 1.5%, less than 2%, less than 2.5%, less than 3%, less than 3.5%, less than 4%, less than 4.5%, less than 5%, less than 6%, less than 7%, less than 8%, less than 9%, less than 10%, less than 12%, less than 15%, or less than 20%. The number of indels formed at a nucleic acid region may depend on the amount of time a nucleic acid *(e.g.,* a nucleic acid within the genome of a cell) is exposed to a base editor. In some embodiments, an number or proportion of indels is determined after at least 1 hour, at least 2 hours, at least 6 hours, at least 12 hours, at least 24 hours, at least 36 hours, at least 48 hours, at least 3 days, at least 4 days, at least 5 days, at least 7 days, at least 10 days, or at least 14 days of exposing a nucleic acid *(e.g.,* a nucleic acid within the genome of a cell) to a base editor.

[00377] Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of efficiently generating an intended mutation, such as a point mutation, in a nucleic acid *(e.g.* a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations, such as unintended point mutations. In some embodiments, a intended mutation is a mutation that is generated by a specific base editor bound to a gRNA, specifically designed to generate the intended mutation. In some embodiments, the intended mutation is a mutation associated with a disease or disorder. In some embodiments, the intended mutation is a cytosine (C) to thymine (T) point mutation associated with a disease or disorder. In some embodiments, the intended mutation is a guanine (G) to adenine (A) point mutation associated with a disease or disorder. In some embodiments, the intended mutation is a cytosine (C) to thymine (T) point mutation within the coding region of a gene. In some embodiments, the intended mutation is a guanine (G) to adenine (A) point mutation within the coding region of a gene. In some embodiments, the intended mutation is a point mutation that generates a stop codon, for example, a premature stop codon within the coding region of a gene. In some embodiments, the intended mutation

is a mutation that eliminates a stop codon. In some embodiments, the intended mutation is a mutation that alters the splicing of a gene. In some embodiments, the intended mutation is a mutation that alters the regulatory sequence of a gene *(e.g.,* a gene promotor or gene repressor). In some embodiments, any of the base editors provided herein are capable of generating a ratio of intended mutations to unintended mutations *(e.g.,* intended point mutations:unintended point mutations) that is greater than 1:1. In some embodiments, any of the base editors provided herein are capable of generating a ratio of intended mutations to unintended mutations *(e.g.,* intended point mutations:unintended point mutations) that is at least 1.5:1, at least 2:1, at least 2.5:1, at least 3:1, at least 3.5:1, at least 4:1, at least 4.5:1, at least 5:1, at least 5.5:1, at least 6:1, at least 6.5:1, at least 7:1, at least 7.5:1, at least 8:1, at least 10:1, at least 12:1, at least 15:1, at least 20:1, at least 25:1, at least 30:1, at least 40:1, at least 50:1, at least 100:1, at least 150:1, at least 200:1, at least 250:1, at least 500:1, or at least 1000:1, or more. It should be appreciated that the characteristics of the base editors described in the *"Base Editor Efficiency"* section, herein, may be applied to any of the fusion proteins, or methods of using the fusion proteins provided herein.

Methods for Editing Nucleic Acids

[00378] Some aspects of the disclosure provide methods for editing a nucleic acid. In some embodiments, the method is a method for editing a nucleobase of a nucleic acid *(e.g.,* a base pair of a double-stranded DNA sequence). In some embodiments, the method comprises the steps of: a) contacting a target region of a nucleic acid *(e.g.,* a double-stranded DNA sequence) with a complex comprising a base editor *(e.g.,* a Cas9 domain fused to a cytidine deaminase domain) and a guide nucleic acid *(e.g.,* gRNA), wherein the target region comprises a targeted nucleobase pair, b) inducing strand separation of said target region, c)converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase, and d) cutting no more than one strand of said target region, where a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase; and the method results in less than 20% indel formation in the nucleic acid. It should be appreciated that in some embodiments, step b is omitted. In some embodiments, the first nucleobase is a cytosine. In some embodiments, the second nucleobase is a deaminated cytosine, or a uracil. In some embodiments, the third nucleobase is a guanine. In some embodiments, the fourth nucleobase is an adenine. In some embodiments, the first nucleobase is a cytosine, the second nucleobase is a deaminated cytosine, or a uracil, the third nucleobase is a guanine, and the fourth nucleobase is an

adenine. In some embodiments, the method results in less than 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2%, or less than 0.1% indel formation. In some embodiments, the method further comprises replacing the second nucleobase with a fifth nucleobase that is complementary to the fourth nucleobase, thereby generating an intended edited base pair *(e.g.,* C:G -> T:A). In some embodiments, the fifth nucleobase is a thymine. In some embodiments, at least 5% of the intended basepaires are edited. In some embodiments, at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, or 50% of the intended basepaires are edited.

[00379] In some embodiments, the ratio of intended products to unintended products in the target nucleotide is at least 2:1, 5:1, 10: 1, 20: 1, 30: 1, 40: 1, 50: 1, 60: 1, 70: 1, 80: 1, 90: 1, 100: 1, or 200: 1, or more. In some embodiments, the ratio of intended point mutation to indel formation is greater than 1:1, 10:1, 50:1, 100:1, 500:1, or 1000:1, or more. In some embodiments, the cut single strand (nicked strand) is hybridized to the guide nucleic acid. In some embodiments, the cut single strand is opposite to the strand comprising the first nucleobase. In some embodiments, the base editor comprises a Cas9 domain. In some embodiments, the first base is cytosine, and the second base is not a G, C, A, or T. In some embodiments, the second base is uracil. In some embodiments, the first base is cytosine. In some embodiments, the second base is not a G, C, A, or T. In some embodiments, the second base is uracil. In some embodiments, the base editor inhibits base escision repair of the edited strand. In some embodiments, the base editor protects or binds the non-edited strand. In some embodiments, the base editor comprises UGI activity. In some embodiments, the base editor comprises nickase activity. In some embodiments, the intended edited basepair is upstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream of the PAM site. In some embodiments, the intended edited basepair is downstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides downstream stream of the PAM site. In some embodiments, the method does not require a canonical *(e.g.,* NGG) PAM site. In some embodiments, the nucleobase editor comprises a linker. In some embodiments, the linker is 1-25 amino acids in length. In some embodiments, the linker is 5-20 amino acids in length. In some embodiments, linker is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length. In some embodiments, the target region comprises a target window, wherein the target window comprises the target nucleobase pair. In some embodiments, the target window comprises 1-10 nucleotides. In some embodiments, the target window is 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-

2, or 1 nucleotides in length. In some embodiments, the target window is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some embodiments, the intended edited base pair is within the target window. In some embodiments, the target window comprises the intended edited base pair. In some embodiments, the method is performed using any of the base editors provided herein. In some embodiments, a target windo is a deamination window

[00380] In some embodiments, the disclosure provides methods for editing a nucleotide. In some embodiments, the disclosure provides a method for editing a nucleobase pair of a double-stranded DNA sequence. In some embodiments, the method comprises a) contacting a target region of the double-stranded DNA sequence with a complex comprising a base editor and a guide nucleic acid *(e.g.,* gRNA), where the target region comprises a target nucleobase pair, b) inducing strand separation of said target region, c) converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase, d) cutting no more than one strand of said target region, wherein a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase, and the second nucleobase is replaced with a fifth nucleobase that is complementary to the fourth nucleobase, thereby generating an intended edited basepair, wherein the efficiency of generating the intended edited basepair is at least 5%. It should be appreciated that in some embodiments, step b is omitted. In some embodiments, at least 5% of the intended basepaires are edited. In some embodiments, at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, or 50% of the intended basepaires are edited. In some embodiments, the method causes less than 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2%, or less than 0.1% indel formation. In some embodiments, the ratio of intended product to unintended products at the target nucleotide is at least 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, or 200:1, or more. In some embodiments, the ratio of intended point mutation to indel formation is greater than 1:1, 10:1, 50:1, 100:1, 500:1, or 1000:1, or more. In some embodiments, the cut single strand is hybridized to the guide nucleic acid. In some embodiments, the cut single strand is opposite to the strand comprising the first nucleobase. In some embodiments, the first base is cytosine. In some embodiments, the second nucleobase is not G, C, A, or T. In some embodiments, the second base is uracil. In some embodiments, the base editor inhibits base escision repair of the edited strand. In some embodiments, the base editor protects or binds the non-edited strand. In some embodiments, the nucleobase editor comprises UGI activity. In some embodiments, the nucleobase edit comprises nickase activity. In some embodiments, the intended edited basepair is upstream of

a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream of the PAM site. In some embodiments, the intended edited basepair is downstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides downstream stream of the PAM site. In some embodiments, the method does not require a canonical (e.g., NGG) PAM site. In some embodiments, the nucleobase editor comprises a linker. In some embodiments, the linker is 1-25 amino acids in length. In some embodiments, the linker is 5-20 amino acids in length. In some embodiments, the linker is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length. In some embodiments, the target region comprises a target window, wherein the target window comprises the target nucleobase pair. In some embodiments, the target window comprises 1-10 nucleotides. In some embodiments, the target window is 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2, or 1 nucleotides in length. In some embodiments, the target window is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some embodiments, the intended edited base pair occurs within the target window. In some embodiments, the target window comprises the intended edited base pair. In some embodiments, the nucleobase editor is any one of the base editors provided herein.

Pharmaceutical compositions

[00381] In some embodiments, any of the fusion proteins, gRNAs, and/or complexes described herein are provided as part of a pharmaceutical composition. In some embodiments, the pharmaceutical composition comprises any of the fusion proteins provided herein. In some embodiments, the pharmaceutical composition comprises any of the complexes provided herein. In some embodiments, the pharmaceutical composition comprises a ribonucleoprotein complex comprising an RNA-guided nuclease *{e.g.,* Cas9) that forms a complex with a gRNA and a cationic lipid. In some embodiments pharmaceutical composition comprises a gRNA, a nucleic acid programmable DNA binding protein, a cationic lipid, and a pharmaceutically acceptable excipient. Pharmaceutical compositions may optionally comprise one or more additional therapeutically active substances.

[00382] In some embodiments, compositions provided herein are administered to a subject, for example, to a human subject, in order to effect a targeted genomic modification within the subject. In some embodiments, cells are obtained from the subject and contacted with a any of the pharmaceutical compositions provided herein. In some embodiments, cells removed from a subject and contacted *ex vivo* with a pharmaceutical composition are re-introduced into the

subject, optionally after the desired genomic modification has been effected or detected in the cells. Methods of delivering pharmaceutical compositions comprising nucleases are known, and are described, for example, in U.S. Pat. Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties. Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals or organisms of all sorts. Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, domesticated animals, pets, and commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as chickens, ducks, geese, and/or turkeys.

[00383] Formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient(s) into association with an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit.

[00384] Pharmaceutical formulations may additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage form desired. Remington's *The Science and Practice of Pharmacy*, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated in its entirety herein by reference) discloses various excipients used in formulating pharmaceutical compositions and known techniques for the preparation thereof. See also PCT application PCT/US2010/055131 (Publication number WO2011053982 A8, filed Nov. 2, 2010), incorporated in its entirety herein by reference, for additional suitable methods, reagents, excipients and solvents for producing pharmaceutical compositions comprising a nuclease. Except insofar as any conventional excipient medium is incompatible

with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition, its use is contemplated to be within the scope of this disclosure.

[00385] In some embodiments, compositions in accordance with the present invention may be used for treatment of any of a variety of diseases, disorders, and/or conditions, including but not limited to one or more of the following: autoimmune disorders *(e.g.* diabetes, lupus, multiple sclerosis, psoriasis, rheumatoid arthritis); inflammatory disorders *(e.g.* arthritis, pelvic inflammatory disease); infectious diseases *(e.g.* viral infections *(e.g.,* HIV, HCV, RSV), bacterial infections, fungal infections, sepsis); neurological disorders *(e.g.* Alzheimer's disease, Huntington's disease; autism; Duchenne muscular dystrophy); cardiovascular disorders *(e.g.* atherosclerosis, hypercholesterolemia, thrombosis, clotting disorders, angiogenic disorders such as macular degeneration); proliferative disorders *(e.g.* cancer, benign neoplasms); respiratory disorders *(e.g.* chronic obstructive pulmonary disease); digestive disorders *(e.g.* inflammatory bowel disease, ulcers); musculoskeletal disorders *(e.g.* fibromyalgia, arthritis); endocrine, metabolic, and nutritional disorders *(e.g.* diabetes, osteoporosis); urological disorders *(e.g.* renal disease); psychological disorders *(e.g.* depression, schizophrenia); skin disorders *(e.g.* wounds, eczema); blood and lymphatic disorders *(e.g.* anemia, hemophilia); *etc.*

Kits, vectors, cells

[00386] Some aspects of this disclosure provide kits comprising a nucleic acid construct, comprising (a) a nucleotide sequence encoding a Cas9 protein or a Cas9 fusion protein as provided herein; and (b) a heterologous promoter that drives expression of the sequence of (a). In some embodiments, the kit further comprises an expression construct encoding a guide RNA backbone, wherein the construct comprises a cloning site positioned to allow the cloning of a nucleic acid sequence identical or complementary to a target sequence into the guide RNA backbone.

[00387] Some aspects of this disclosure provide polynucleotides encoding a Cas9 protein of a fusion protein as provided herein. Some aspects of this disclosure provide vectors comprising such polynucleotides. In some embodiments, the vector comprises a heterologous promoter driving expression of polynucleotide.

[00388] Some aspects of this disclosure provide cells comprising a Cas9 protein, a fusion protein, a nucleic acid molecule encoding the fusion protein, a complex comprise the Cas9 protein and the gRNA, and/or a vector as provided herein.

[00389] The description of exemplary embodiments of the reporter systems above is provided for illustration purposes only and not meant to be limiting. Additional reporter systems, *e.g.,* variations of the exemplary systems described in detail above, are also embraced by this disclosure.

EXAMPLES

EXAMPLE 1: Cas9 Deaminase Fusion Proteins

[00390] A number of Cas9:Deaminase fusion proteins were generated and deaminase activity of the generated fusions was characterized. The following deaminases were tested:

Human AID (hAID):
MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPYLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS
VRLSRQLRRILLPLYEVDDLRDAFRTLGLLD (SEQ ID NO: 49)

Human AID-DC (hAID-DC, truncated version of hAID with 7-fold increased activity):
MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS
VRLSRQLRRILL (SEQ ID NO: 50)

Rat APOBEC1 (rAPOBECl):
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ ID
NO: 76)

Human APOBEC1 (hAPOBECl)
MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNT
TNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVA
RLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPP
LWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIH
PSVAWR (SEQ ID NO: 74)

Petromyzon marinus (Lamprey) CDA1 (pmCDAl):
MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKP
QSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGN
GHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQL
NENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO: 81)

Human APOBEC3G (hAPOBEC3G):
MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVT
LTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKTMNYDEFQHCWSKFVYSQRELF
EPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHN

DTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSW
SPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFK
HCWDTFVDHQGCPFQPWDGLDEHS QDLS GRLRAILQNQEN (SEQ ID NO: 60)

[00391] Deaminase Activity on ssDNA. A USER (Uracil-Specific Excision Reagent) Enzyme-based assay for deamination was employed to test the activity of various deaminases on single-stranded DNA (ssDNA) substrates. USER Enzyme was obtained from New England Biolabs. An ssDNA substrate was provided with a target cytosine residue at different positions. Deamination of the ssDNA cytosine target residue results in conversion of the target cytosine to a uracil. The USER Enzyme excises the uracil base and cleaves the ssDNA backbone at that position, cutting the ssDNA substrate into two shorter fragments of DNA. In some assays, the ssDNA substrate is labeled on one end with a dye, *e.g.*, with a 5' Cy3 label (the * in the scheme below). Upon deamination, excision, and cleavage of the strand, the substrate can be subjected to electrophoresis, and the substrate and any fragment released from it can be visualized by detecting the label. Where Cy5 is images, only the fragment with the label will be visible via imaging.

[00392] In one USER Enzyme assay, ssDNA substrates were used that matched the target sequences of the various deaminases tested. Expression cassettes encoding the deaminases tested were inserted into a CMV backbone plasmid that has been used previously in the lab (Addgene plasmid 52970). The deaminase proteins were expressed using a TNT Quick Coupled Transcription/Translation System (Promega) according to the manufacturers recommendations. After 90 min of incubation, 5 mL of lysate was incubated with 5' Cy3-labeled ssDNA substrate and 1 unit of USER Enzyme (NEB) for 3 hours. The DNA was resolved on a 10% TBE PAGE gel and the DNA was imaged using Cy-dye imaging. A schematic reparesentation of the USER Enzyme assay is shown in Figure 41.

[00393] **Figure 1** shows the deaminase activity of the tested deaminases on ssDNA substrates, such as Doench 1, Doench 2, G7' and VEGF Target 2. The rAPOBECl enzyme exhibited a substantial amount of deamination on the single-stranded DNA substrate with a canonical NGG PAM, but not with a negative control non-canonical NNN PAM.

Cas9 fusion proteins with APOBEC family deaminases were generated. The following fusion architectures were constructed and tested on ssDNA:

rAPQBEC 1-*GGS-dCas9* primary sequence

MSSETGPVAVDPTLRRRIEPHEFE VFFDPRELRKETCLLYEINWGGRHS IWRHTS QNTN
KHVEVNFIEKFTTERYFCPNTRCS IT WFLSWSPCGECSRAITEFLS RYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR

LYVLELYCnLGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKpGSKKKY
SIGLAIGTNSVGWA VITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA TRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYA GYIDGGA SQEEF YKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TP WNFEE W DKGA SA QSFIERMTNFDKNLPNEKVLPKHSLL YE YFTVYNELTKVKYV
TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN
ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT
FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKWDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEW
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA
WGTALIKKYPKLESEFVYGD YKVYD VRKMIAKSEQEIGKA TAKYFF YSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFA TVRKVLSMPQ VNIVKKTE VQ TGGFSKE
SILPKRNSDKLIARKKD WDPKKYGGFDSPTVA YS VL WAKVEKGKSKKLKS VKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 149)

rAPQBEC 1-(***GGSh-dCas9*** primary sequence

MSSETGPVA VDPTLRRRIEPHEFE VFFDPRELRKETCLLYEINWGGRHS IWRHTS QNTN
KHVEVNFIEKFTTERYFCPNTRCS rrWFLSWSPCGECSRAITEFLS RYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK|GGSGGSG|
\GS\MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG
ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD
LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKE
IFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT
RKSEETITPWNFEEWDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDS VEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEWKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKL VSDFRKDFQF YKVREINNYHHAH
DAYLNAWGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM

NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ TGGFSKESILPKRNSDKLIARKKD WDPKKYGGFDSPTVA YS VL WAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF SKR VILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 150)

rfCas9-lGGSl-rAPOBECl

DKKYSIGLAIGTNS VG WA VITDE YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE A TRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA QIGDQ YADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRK SEETITPWNFEEWDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKWDELVKVMGRHKPE NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY LQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VA QILDSRMNTKYDENDKLIRE VKVITLKSKL VSDFRKDFQF YKVREINNYHHAHDA Y LNAWGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG FSKESILPKRNSDKLIA RKKD WDPKKYGGFDSPTVA YS VL WAKVEKGKSKKLKS VKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR YTS TKEVLDATLIHOSITGLYETRIDLSOLGGDpGSMSSKTGPV AY OP*TLRRR*IEPHEFEVF FDPRELRKETCLLYEINWGGRHS IWRHTS QNTNKHVE VNFIEKFTTERYFCPNTRCS ГГ WFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMT EQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQL TFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 151)

dCa***s9***-GGS$_3$-rAPOBEC 1

DKKYSIGLAIGTNS VG WA VITDE YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE A TRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA QIGDQ YADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRK SEETITPWNFEEWDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE

DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKWDELVKVMGRHKPE NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY LQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VA QILDSRMNTKYDENDKLIRE VKVITLKSKL VSDFRKDFQF YKVREINNYHHAHDA Y LNAWGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG FSKESILPKRNSDKLIA RKKD WDPKKYGGFDSPTVA YS VL WAKVEKGKSKKLKS VKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVLSA YNKHRDKPIREQAEN IIHLFTLTNLG APAAFKYFDTTIDRKR YTS TKEVLDA TLIHQSI TGL YE TRIDLSQLGGD GGSGGSGGSGGSIMSSETGPVA VDPTLRRRIEP HEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCP NTRCS rrWFLS WSPCGEC SRAITEFLS RYPH VTLFIYIARLYHH ADPRNRQGLRDLIS SG VTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNIL RRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 152)

rAPQBEC 1-pXTEN-dCas9 primary sequence

MSSETGPVA VDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHS IWRHTS QNTN KHVEVNFIEKFTTERYFCPNTRCS rrWFLS WSPCGEC SRAITEFLS RYPH VTLFIYIARL YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGl TSESATPESp KKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS RFA WMTRKSEE TITP WNFEE W DKGA SA QSFIERMTNFDKNLPNEKVLPKHSLL YE Y FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKWDELVK VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKL YL YYLQNGRDMYVDQELDINRLSD YD VDAIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN NYHHAHDA YLNA WGTALIKKYPKLESEFVYGD YKVYD VRKMIAKSEQEIGKA TAKYF FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLWAKVEKGK SKKLKS VKELLGITIMERSSFEKNPIDFLEAKG YKE VKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKR VILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD* (SEQ ID NO: 153)

[00394] **Figure 2** shows that the N-terminal deaminase fusions showed significant activity on the single stranded DNA substrates. For this reason, only the N-terminal architecture was chosen for further experiments.

[00395] **Figure 3** illustrates double stranded DNA substrate binding by deaminase-dCas9:sgRNA complexes. A number of double stranded deaminase substrate sequences were generated. The sequences are provided below. The structures according to **Figure 3** are identified in these sequences (36bp: underlined, sgRNA target sequence: bold; PAM: boxed; 21bp: italicized). All substrates were labeled with a 5'-Cy3 label:

2:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGTCCCGCGGATTTATTTATTTA
MV3GATGACCTCTGGATCCATGGAC-y (SEQ ID NO: 85)
3:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCTTCCGCGGATTTATTTATTT
MV3GATGACCTCTGGATCCATGGAC-y (SEQ ID NO: 86)
4:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TTCCGCGGATTTATTTATT**
ATGG*ATGACCTCTGGATCCATGGAC*-3' (SEQ ID NO: 87)
5:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTCCGCGGATTTATTTAT
l^GGATGACCTCTGGATCCATGGAC-y (SEQ ID NO: 88)
6:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTATTCCGCGGATTTATTTA
T\TGQATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 89)
7:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTTATTCCGCGGATTTATTT
M^GGATGACCTCTGGATCCATGGAC-y (SEQ ID NO: 90)
8:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTCCGCGGATTTATT
TUXKMTGACCTCTGGATCCATGGAC-y (SEQ ID NO: 91)
9:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTATTATTCCGCGGATTTAT
l^GGATGACCTCTGGATCCATGGAC-y (SEQ ID NO: 92)
10:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATATTCCGCGGATTT**
ATTGG*ATGACCTCTGGATCCATGGAC*-3' (SEQ ID NO: 93)
11:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TATTATATTCCGCGGATT**
TATGG*ATGACCTCTGGATCCATGGAC*-3' (SEQ ID NO: 94)
12:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TTATTATATTCCGCGGAT**
TTTGG*ATGACCTCTGGATCCATGGAC*-3' (SEQ ID NO: 95)
13:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATTATATTCCGCGGA**
TTTGG*ATGACCTCTGGATCCATGGAC*-3' (SEQ ID NO: 96)
14:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TATTATTATATTCCGCGG**
ATTGG*ATGACCTCTGGATCCATGGAC*-3' (SEQ ID NO: 97)
15:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATTATTATTACCGCG**
GATGG*ATGACCTCTGGATCCATGGAC*-3' (SEQ ID NO: 98)
18:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTATTATTATTACCG
ʊJGGATGACCTCTGGATCCATGGAC-y (SEQ ID NO: 99)
"-" :
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGTAATATTAATTTATTTATTTAAT
UGATGACCTCTGGATCCATGGAC-y (SEQ ID NO: 100)
8U:GTAGGTAGTTAGGATGAATGGAAGGTTGGTGTAG**ATTATTATCUGCGGATTTAT**
TGG*ATGACCTCTGGATCCATGGACAT*-3' (SEQ ID NO: 101)

*In all substrates except for "8U", the top strand in Figure 3 is the complement of the

sequence specified here. In the case of "8U", there is a "G" opposite the U.

[00396] **Figure 4** shows the results of a double stranded DNA Deamination Assay. The fusions were expressed and purified with an N-terminal His6 tag via both Ni-NTA and sepharose chromatography. In order to assess deamination on dsDNA substrates, the various dsDNA substrates shown on the previous slide were incubated at a 1:8 dsDNA:fusion protein ratio and incubated at 37 ℃ for 2 hours. Once the dCas9 portion of the fusion binds to the DNA it blocks access of the USER enzyme to the DNA. Therefore, the fusion proteins were denatured following the incubation and the dsDNA was purified on a spin column, followed by incubation for 45 min with the USER Enzyme and resolution of the resulting DNA substrate and substrate fragments on a 10% TBE-urea gel.

[00397] **Figure 5** demonstrates that Cas9 fusions can target positions 3-11 of double-stranded DNA target sequences (numbered according to the schematic in Figure 3). Upper Gel: 1 μM rAPOBECl-GGS-dCas9, 125 nM dsDNA, 1 eq sgRNA. Mid Gel: 1 μM rAPOBECl-$(GGS)_3$-dCas9, 125 nM dsDNA, 1 eq sgRNA. Lower Gel: 1.85 μM rAPOBECl-XTEN-dCas9, 125 nM dsDNA, 1 eq sgRNA. Based on the data from these gels, positions 3-11 (according to the numbering in Figure 3) are sufficiently exposed to the activity of the deaminase to be targeted by the fusion proteins tested. Access of the deaminase to other positions is most likely blocked by the dCas9 protein.

[00398] The data further indicates that a linker of only 3 amino acids (GGS) is not optimal for allowing the deaminase to access the single stranded portion of the DNA. The 9 amino acid linker $[(GGS)_3]$ (SEQ ID NO: 610) and the more structured 16 amino acid linker (XTEN) allow for more efficient deamination.

[00399] **Figure 6** demonstrates that the correct guide RNA, *e.g.,* the correct sgRNA, is required for deaminase activity. The gel shows that fusing the deaminase to dCas9, the deaminase enzyme becomes sequence specific *(e.g.,* using the fusion with an eGFP sgRNA results in no deamination), and also confers the capacity to the deaminase to deaminate dsDNA. The native substrate of the deaminase enzyme is ssDNA, and no deamination occurred when no sgRNA was added. This is consistent with reported knowledge that APOBEC deaminase by itself does not deaminate dsDNA. The data indicates that Cas9 opens the double-stranded DNA helix within a short window, exposing single-stranded DNA that is then accessible to the APOBEC deaminase for cytidine deamination. The sgRNA sequences used are provided below. sequences (36bp: underlined, sgRNA target sequence: bold; PAM: boxed; 21bp: italicized)

DNA sequence 8:

5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTCCGCGGATTTATT
ITiGGATGACCTCTGGATCCATGGAC-y (SEQ ID NO: 102)

Correct sgRNA sequence (partial 3' sequence):
5'-AUUAUUCCGCGGAUUUAUUUGUUUUAGAGCUAG...-3' (SEQ ID NO: 103)

eGFP sgRNA sequence (partial 3'-sequence):
5'-CGUAGGCCAGGGUGGUCACGGUUUUAGAGCUAG ...-**3'** (SEQ ID NO: 104)

EXAMPLE 2: Deamination of DNA target sequence

[00400] Exemplary deamination targets. The dCas9:deaminase fusion proteins described herein can be delivered to a cell in vitro or ex vivo or to a subject *in vivo* and can be used to effect C to T or G to A transitions when the target nucleotide is in positions 3-11 with respect to a PAM. Exemplary deamination targets include, without limitation, the following: CCR5 truncations: any of the codons encoding **Q93, Q102, Q186, R225, W86,** or **Q261** of CCR5 can be deaminated to generate a STOP codon, which results in a nonfunctional truncation of CCR5 with applications in HIV treatment. APOE4 mutations: mutant codons encoding **C11R** and **C57R** mutant APOE4 proteins can be deaminated to revert to the wild-type amino acid with applications in Alzheimer's treatment. eGFP truncations: any of the codons encoding **Q158, Q184, Q185** can be deaminated to generate a STOP codon, or the codon encoding **Ml** can be deaminated to encode I, all of which result in loss of eGFP fluorescence, with applications in reporter systems. eGFP restoration: a mutant codon encoding **T65A** or **Y66C** mutant GFP, which does not exhibit substantial fluorescence, can be deaminated to restore the wild-type amino acid and confer fluorescence. PIK3CA mutation: a mutant codon encoding **K111E** mutant PIK3CA can be deaminated to restore the wild-type amino acid residue with applications in cancer. CTNNB 1 mutation: a mutant codon encoding **T41A** mutant CTNNB 1 can be deaminated to restore the wild-type amino acid residue with applications in cancer. HRAS mutation: a mutant codon encoding **Q61R** mutant HRAS can be deaminated to restore the wild-type amino acid residue with applications in cancer. P53 mutations: any of the mutant codons encoding **Y163C, Y236C,** or **N239D** mutant p53 can be deaminated to encode the wild type amino acid sequence with applications in cancer.

The feasibility of deaminating these target sequences in double-stranded DNA is demonstrated in **Figures 7** and **8. Figure 7** illustrates the mechanism of target DNA binding of *in vivo* target sequences by deaminase-dCas9:sgRNA complexes.

[00401] **Figure 8** shows successful deamination of exemplary disease-associated target sequences. Upper Gel: CCR5 Q93: coding strand target in pos. 10 (potential off-targets at positions 2, **5,** 6, 8, 9); CCR5 Q102: coding strand target in pos. **9** (potential off-targets at positions 1, 12, 14); CCR5 Q186: coding strand target in pos. 9 (potential off-targets at positions 1, **5,** 15); CCR5 R225: coding strand target in pos. **6** (no potential off-targets); eGFP Q158: coding strand target in pos. **5** (potential off-targets at positions 1, 13, 16); eGFP Q184 /185: coding strand target in pos. **4** and 7 (potential off-targets at positions 3, 12, 14, 15, 16, 17, 18); eGFP M l : template strand target in pos. 12 (potential off-targets at positions 2, **3,** 7, 9, 11) (targets positions 7 and 9 to small degree); eGFP T65A: template strand target in pos. 7 (potential off-targets at positions 1, **8,** 17); PIK3CAK111E: template strand target in pos. 2 (potential off-targets at positions **5, 8,** 10, 16, 17); PIK3CAK111E: template strand target in pos. 13 (potential off-targets at positions 11, 16, 19) X. Lower Gel: CCR5 W86: template strand target in pos. 2 and 3 (potential off-targets at positions 1, 13) X; APOE4 CllR: coding strand target in pos. 11 (potential off-targets at positions **7,** 13, 16, 17); APOE4 C57R: coding strand target in pos. **5)** (potential off-targets at positions 7, 8, 12); eGFP Y66C: template strand target in pos. 11 (potential off-targets at positions 1, 4 **,6** , 8, 9, 16); eGFP Y66C: template strand target in pos. 3 (potential off-targets at positions 1, **8,** 17); CCR5 Q261: coding strand target in pos. 10 (potential off-targets at positions 3, **5, 6,** 9, 18); CTNNB 1 T41A: template strand target in pos. 7 (potential off-targets at positions 1, 13, 15, 16) X; HRAS Q61R: template strand target in pos. 6 (potential off-targets at positions 1, 2, 4, **5,** 9, 10, 13); p53 Y163C: template strand target in pos. **6** (potential off-targets at positions 2, 13, 14); p53 Y236C: template strand target in pos. 8 (potential off-targets at positions 2, **4);** p53 N239D: template strand target in pos. **4** (potential off-targets at positions 6, 8). Exemplary DNA sequences of disease targets are provided below (PAMs (5'-NGG-3') and target positions are boxed):

CCR5 **Q93:** 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAACTATGCTGCCGCC
C|A*G*T*GGG*A|C*T*T|T*GG*|A*AATACAATGTGTCAACTCTT-y* (SEQ **ID NO:** 105)
CCR5 **Q102:** 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAAATACAATGTGT
C|AACTCTT*G*ACA|GGG|*CTCTATTTTATAGGCTTCTTC*-3' (SEQ **ID NO:** 106)
CCR5 **Q186:** 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATTTTCCATACAGT
[C|AGTATCAATTC |TGG|AA*GAATTTCCAGACATTAAAG*-3' (SEQ **ID NO:** 107)
CCR5 **R225:** 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGCTTCGGTGTgGA
AATGA GAA GAA G|AGG|CA*CA GGGCTGTGA GGCTTATC*-3' (SEQ **ID NO:** 108)

CCR5 **W86:** 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGTGAGCggAGAAGG
GGAC A GTA A GA|AGG|AAAAA *CA GGTCAGA GATGGCC-3*' (SEQ **ID** NO: 109)
CCR5 **Q261:** 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATCCTG**AACACCTT**
C|C|AGGAATTCTT|TGG|*CCTGAATAATTGCAGTAGCTC*-3' (SEQ ID NO: 110)
APOE4 **C11R**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGACAT**GGAGGAC**
GTG|C|GCGGCCGCC|TGG|*TGCAGTACCGCGGCGAGGTGC*-3' (SEQ ID NO: 111)
APOE4 **C57R**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTGCA**GAAG|C|GC**
CTGGCAGTGTACC|AGG|*CCGGGGCCCGCGAGGGCGCCG*-3' (SEQ ID NO: 112)
eGFP **Q158**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGCCGA**CAAG|C|AGA**
AGAACGGCATCA|AGG|*TGAACTTCAAGATCCGCCACA*-3' (SEQ ID NO: 113)
eGFP **Q184/185**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAACCAC**TAC|C|**
AG|C|AGAACACCCCCAT|CGG|*CGACGGCCCCGTGCTGCTGCC*-3' (SEQ ID NO: 114)
eGFP **M1**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACCTCG**CCCTTGCTCA**
C|C|ATCTCGAGT|CGG|*CCGCCAGTGTGATGGATATCT*-3' (SEQ ID NO: 115)
eGFP **T65A**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACACGC**GTAGG|C|CA**
GGGTGGTCACG|AGG|*GTGGGCCAGGGCACGGGCAGC*-3' (SEQ ID NO: 116)
eGFP **Y66C**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAAGCA**CTGCACTC**
CG|C|AGGTCAGGG|TGG|*TCACGAGGGTTGGCCAGGGCA*-3' (SEQ ID NO: 117)
eGFP **Y66C**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACACTC**CG|C|AGGTC**
AGGGTGGTCACG|AGG|*GTTGGCCAGGGCACGGGCAGG-3*' (SEQ ID NO: 118)
PIK3CA **K111E**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGATC**T|C|TTC**
TTCACGGTTGCCTAC|TGG|*TTCAATTACTTTTAAAAATGG*-3' (SEQ ID NO: 119)
PIK3CA **K111E**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTATTCTC**GATTG**
AGGATCT|C|TTCTTCA|CGG|*TTGCCTACTGGTTCAATTACT*-3' (SEQ ID NO: 120)
CTNNB1 **T41A**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAGGAG**CTGTGG**
|C|AGTGGCACCAGAA|TGG|*ATTCCAGAGTCCAGGTAAGAC*-3' (SEQ ID NO: 121)
HRAS **Q61R**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGTACT**CCTCC|C|GG**
CCGGCGGTATCC|AGG|*ATGTCCAACAGGCACGTCTCC*-3' (SEQ ID NO: 122)
p53 **Y163C:** 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATGACTGCTTGgAG
ATGGCCATGGCG |**CGG**|ACGCGGGrGCCGGGCGGGGGr-3' (SEQ **ID** NO: 123)
p53 **Y236C:** 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTGTTACACATGg
AGTTGTAGTGGA ⌈**rGG**|*TGGTA*CAGrCAGAGCCAACCr-3' (SEQ **ID** NO: 124)
p53 **N239D:** 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGAACTGTgACAC
ATGTAGTTGTAG |**TGG**|ArGGrGG7ACAGrCAGAGCCA-3' (SEQ **ID** NO: 125)

EXAMPLE 3: Uracil Glycosylase Inhibitor Fusion Improves Deamination Efficiency

[00402] Direct programmable nucleobase editing efficiencies in mammalian cells by dCas9:deaminase fusion proteins can be improved significantly by fusing a uracil glycosylase inhibitor (UGI) to the dCas9:deaminase fusion protein.

[00403] **Figure 9** shows *in vitro* C→T editing efficiencies in human HEK293 cells using rAPOBECl-XTEN-dCas9:

rAPOBEC 1-XTEN-dCas9 |-*NLS* primary sequence

MSSETGPVAVDPTLRRRIEPHEFE VFFDPRELRKETCLLYEINWGGRHS IWRHTS QNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTS
ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
pSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKl
|HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
|GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPl
|GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAl
pLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK|
|YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFl
pNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMl
|TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE|
|LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
|GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
|HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI]
|HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
|KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLl
|YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
|PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIj
|TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH|
pAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI]
|MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVl
|QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLl
|KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
|AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
|EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRl
|KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD|SGGSP^^^y (SEQ ID NO: 126)

Protospacer sequences were as follows:

EMX1: 5'- GAGTC_5C_6GAGC_{10}AGAAGAAGAAGGG -3' (SEQ ID NO: 127)

FANCF: 5'- GGAAT$C_6C_7C_8$TTC_{11}TGCAGCACCTGG -3' (SEQ ID NO: 128)

HEK293 site 2: 5'- GAAC_4AC_6AAAGC_{11}ATAGACTGCGGG -3' (SEQ ID NO: 129)

HEK293 site 3: 5'- GG$C_3C_4C_5$AGAC_9TGAGCACGTGAATGG -3' (SEQ ID NO: 130)

HEK293 site 4: 5'- GGC$_3$AC$_5$TGC$_8$GGC$_{11}$TGGAGGTGG |GGG| -3' (SEQ ID NO: 735)
RNF2: 5'- GTC$_3$ATC$_6$TTAGTCATTACCTG |AGG| -3' (SEQ ID NO: 132)
*PAMs are boxed, C residues within target window (positions 3-11) are numbered and bolded.

[00404] **Figure 10** demonstrates that C→T editing efficiencies on the same protospacer sequences in HEK293T cells are greatly enhanced when a UGI domain is fused to the rAPOBECl:dCas9 fusion protein.

rAPOBEC 1-**XTEN**-^Cas9j-**UGI**-*NLS* primary sequence

MSSETGPVAVDPTLRRRIEPHEFE VFFDPRELRKETCLLYEINWGGRHS IWRHTS QNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTS
ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
PSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKI
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPl
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAl
pLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFl
pNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMl
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI]
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYI]
YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIj
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
pAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI]
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVl
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLl
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRl
KRYTSTKE VLDATLIHQS ITGLYETRIDLS QLGGD|S GGS **TNLSDIIEKETGKQLVIQESI**
LMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGE
NKIKMLS *GGSPKKKRKV* (SEQ ID NO: 133)

[00405] The percentages in Figures 9 and 10 are shown from sequencing both strands of the target sequence. Because only one of the strands is a substrate for deamination, the maximum possible deamination value in this assay is 50%. Accordingly, the deamination efficiency is

double the percentages shown in the tables. *E.g.,* a value of 50% relates to deamination of 100% of double-stranded target sequences.

When a uracil glycosylase inhibitor (UGI) was fused to the dCas9:deaminase fusion protein *(e.g.,* rAPOBECl-XTEN-dCas9-[UGI]-NLS), a significant increase in editing efficiency in cells was observed. This result indicates that in mammalian cells, the DNA repair machinery that cuts out the uracil base in a U:G base pair is a rate-limiting process in DNA editing. Tethering UGI to the dVas9deaminase fusion proteins greatly increases editing yields.

[00406] Without UGI, typical editing efficiencies in human cells were in the -2-14% yield range (Figure 9 and Figure 10, "XTEN" entries). With UGI (Figure 10, "UGI" entries) the editing was observed in the -6-40% range. Using a UGI fusion is thus more efficient than the current alternative method of correcting point mutations via HDR, which also creates an excess of indels in addition to correcting the point mutation. No indels resulting from treatment with the cas9:deaminase:UGI fusions were observed.

EXAMPLE 4: Direct, programmable conversion of a target nucleotide in genomic DNA without double-stranded DNA cleavage

[00407] Current genome-editing technologies introduce double-stranded DNA breaks at a target locus of interest as the first step to gene correction.[39,40] Although most genetic diseases arise from mutation of a single nucleobase to a different nucleobase, current approaches to revert such changes are very inefficient and typically induce an abundance of random insertions and deletions (indels) at the target locus as a consequence of the cellular response to double-stranded DNA breaks.[39,40] Reported herein is the development of nucleobase editing, a new strategy for genome editing that enables the direct conversion of one target nucleobase into another in a programmable manner, without requiring double-stranded DNA backbone cleavage. Fusions of CRISPR/Cas9 were engineered and the cytidine deaminase enzyme APOBEC1 that retain the ability to be programmed with a guide RNA, do not induce double-stranded DNA breaks, and mediate the direct conversion of cytidine to uracil, thereby effecting a C→T (or G→A) substitution following DNA replication, DNA repair, or transcription if the template strand is targeted. The resulting "nucleobase editors" convert cytidines within a window of approximately five nucleotides, and can efficiently correct a variety of point mutations relevant to human disease *in vitro.* In four transformed human and murine cell lines, second- and third-generation nucleobase editors that fuse uracil glycosylase inhibitor (UGI), and that use a Cas9 nickase targeting the non-edited strand, respectively, can

overcome the cellular DNA repair response to nucleobase editing, resulting in permanent correction of up to 37% or (-15-75%) of total cellular DNA in human cells with minimal (typically ≤ 1%) indel formation. In contrast, canonical Cas9-mediated HDR on the same targets yielded an average of 0.7% correction with 4% indel formation. Nucleobase editors were used to revert two oncogenic *p53* mutations into wild-type alleles in human breast cancer and lymphoma cells, and to convert an Alzheimer's Disease associated Arg codon in *ApoE4* into a non-disease-associated Cys codon in mouse astrocytes. Base editing expands the scope and efficiency of genome editing of point mutations.

[00408] The clustered regularly interspaced short palindromic repeat (CRISPR) system is a prokaryotic adaptive immune system that has been adapted to mediate genome engineering in a variety of organisms and cell lines.[41] CRISPR/Cas9 protein-RNA complexes localize to a target DNA sequence through base pairing with a guide RNA, and natively create a DNA double-stranded break (DSB) at the locus specified by the guide RNA. In response to DSBs, endogenous DNA repair processes mostly result in random insertions or deletions (indels) at the site of DNA cleavage through non-homologous end joining (NHEJ). In the presence of a homologous DNA template, the DNA surrounding the cleavage site can be replaced through homology-directed repair (HDR). When simple disruption of a disease-associated gene is sufficient (for example, to treat some gain-of-function diseases), targeted DNA cleavage followed by indel formation can be effective. For most known genetic diseases, however, correction of a point mutation in the target locus, rather than stochastic disruption of the gene, is needed to address or study the underlying cause of the disease.[68]

[00409] Motivated by this need, researchers have invested intense effort to increase the efficiency of HDR and suppress NHEJ. For example, a small-molecule inhibitor of ligase IV, an essential enzyme in the NHEJ pathway, has been shown to increase HDR efficiency.[42,43] However, this strategy is challenging in post-mitotic cells, which typically down-regulate HDR, and its therapeutic relevance is limited by the potential risks of inhibiting ligase IV in non-target cells. Enhanced HDR efficiency can also be achieved by the timed delivery of Cas9-guide RNA complexes into chemically synchronized cells, as HDR efficiency is highly cell-cycle dependent.[44] Such an approach, however, is limited to research applications in cell culture since synchronizing cells is highly disruptive. Despite these developments, current strategies to replace point mutations using HDR in most contexts are very inefficient (typically -0.1 to 5%),[42,43,45,46,75] especially in unmodified, non-dividing cells. In addition, HDR competes with NHEJ during the resolution of double-stranded breaks, and indels are generally more abundant outcomes than gene replacement. These observations highlight the need to

develop alternative approaches to install specific modifications in genomic DNA that do not rely on creating double-stranded DNA breaks. A small-molecule inhibitor of ligase IV, an essential enzyme in the NHEJ pathway, has been shown to increase HDR efficiency.[42,43] However, this strategy is challenging in post-mitotic cells, which typically down-regulate HDR, and its therapeutic relevance is limited by the potential risks of inhibiting ligase IV in non-target cells. Enhanced HDR efficiency can also be achieved by the timed delivery of Cas9-guide RNA complexes into chemically synchronized cells, as HDR efficiency is highly cell-cycle dependent.[44] Such an approach, however, is limited to research applications in cell culture since synchronizing cells is highly disruptive. In some cases, it is possible to design HDR templates such that the product of successful HDR contains mutations in the PAM sequence and therefore is no longer a substrate for subsequent Cas9 modification, increasing the overall yield of HDR products,[75] although such an approach imposes constraints on the product sequences. Recently, this strategy has been coupled to the use of ssDNA donors that are complementary to the non-target strand and high-efficiency ribonucleoprotein (RNP) delivery to substantially increase the efficiency of HDR, but even in these cases the ratio of HDR to NHEJ outcomes is relatively low (< 2).[83]

[00410] It was envisioned that direct catalysis of the conversion of one nucleobase to another at a programmable target locus without requiring DNA backbone cleavage could increase the efficiency of gene correction relative to HDR without introducing undesired random indels at the locus of interest. Catalytically dead Cas9 (dCas9), which contains AsplOAla and His840Ala mutations that inactivate its nuclease activity, retains its ability to bind DNA in a guide RNA-programmed manner but does not cleave the DNA backbone.[16,47] In principle, conjugation of dCas9 with an enzymatic or chemical catalyst that mediates the direct conversion of one nucleobase to another could enable RNA-programmed nucleobase editing. The deamination of cytosine (C) is catalyzed by cytidine deaminases[29] and results in uracil (U), which has the base pairing properties of thymine (T). dCas9 was fused to cytidine deaminase enzymes in order to test their ability to convert C to U at a guide RNA-specified DNA locus. Most known cytidine deaminases operate on RNA, and the few examples that are known to accept DNA require single-stranded DNA.[48] Recent studies on the dCas9-target DNA complex reveal that at least nine nucleotides of the displaced DNA strand are unpaired upon formation of the Cas9:guide RNA:DNA "R-loop" complex.[12] Indeed, in the structure of the Cas9 R-loop complex the first 11 nucleotides of the protospacer on the displaced DNA strand are disordered, suggesting that their movement is not highly restricted.[76] It has also been speculated that Cas9 nickase-induced mutations at cytosines in the non-template strand might

arise from their accessibility by cellular cytidine deaminase enzymes.[77] Recent studies on the dCas9-target DNA complex have revealed that at least 26 bases on the non-template strand are unpaired when Cas9 binds to its target DNA sequence.[49] It was reasoned that a subset of this stretch of single-stranded DNA in the R-loop might serve as a substrate for a dCas9-tethered cytidine deaminase to effect direct, programmable conversion of C to U in DNA (Figure 11A).

[00411] Four different cytidine deaminase enzymes (hAID, hAPOBEC3G, rAPOBEC 1, and pmCDAl) were expressed in a mammalian cell lysate-derived *in vitro* transcription-translation system and evaluated for ssDNA deamination. Of the four enzymes, rAPOBEC 1 showed the highest deaminase activity under the tested conditions and was chosen for dCas9 fusion experiments (Figure 36A). Although appending rAPOBEC 1 to the C-terminus of dCas9 abolishes deaminase activity, fusion to the N-terminus of dCas9 preserves deaminase activity on ssDNA at a level comparable to that of the unfused enzyme. Four rAPOBEC l-dCas9 fusions were expressed and purified with linkers of different length and composition (Figure 36B), and evaluated each fusion for single guide RNA (sgRNA)-programmed dsDNA deamination *in vitro* (Figures 11A to 11C and Figures 15A to 15D).

[00412] Efficient, sequence-specific, sgRNA-dependent C to U conversion was observed *in vitro* (Figures 11A to 11C). Conversion efficiency was greatest using rAPOBEC l-dCas9 linkers over nine amino acids in length. The number of positions susceptible to deamination (the deamination "activity window") increases with linker length was extended from three to 2 1 amino acids (Figures 36C to 36F15A to 15D). The 16-residue XTEN linker[50] was found to offer a promising balance between these two characteristics, with an efficient deamination window of approximately five nucleotides, from positions 4 to 8 within the protospacer, counting the end distal to the protospacer-adjacent motif (PAM) as position 1. The rAPOBEC l-XTEN-dCas9 protein served as the first-generation nucleobase editor (NBEl).

[00413] Elected were seven mutations relevant to human disease that in theory could be corrected by C to T nucleobase editing, synthesized double-stranded DNA 80-mers of the corresponding sequences, and assessed the ability of NBEl to correct these mutations *in vitro* (Figures 16A to 16B). NBEl yielded products consistent with efficient editing of the target C, or of at least one C within the activity window when multiple Cs were present, in six of these seven targets *in vitro*, with an average apparent editing efficiency of 44% (Figures 16A to 16B). In the three cases in which multiple Cs were present within the deamination window, evidence of deamination of some or all of these cytosines was observed. In only one of the seven cases tested were substantial yields of edited product observed (Figures 16A to 16B).

Although the preferred sequence context for APOBEC1 substrates is reported to be CC or TC,[51] it was anticipated that the increased effective molarity of the deaminase and its single-stranded DNA substrate mediated by dCas9 binding to the target locus may relax this restriction. To illuminate the sequence context generality of NBEl, its ability to edit a 60-mer double-stranded DNA oligonucleotide containing a single fixed C at position 7 within the protospacer was assayed, as well as all 36 singly mutated variants in which protospacer bases 1-6 and 8-13 were individually varied to each of the other three bases. Each of these 37 sequences were treated with 1.9 μM NBEl, 1.9 μM of the corresponding sgRNA, and 125 nM DNA for 2 h, similar to standard conditions for *in vitro* Cas9 assays[52]. High-throughput DNA sequencing (HTS) revealed 50 to 80% C to U conversion of targeted strands (25 to 40% of total sequence reads arising from both DNA strands, one of which is not a substrate for NBEl) (Figure 12A). The nucleotides surrounding the target C had little effect on editing efficiency was independent of sequence context unless the base immediately 5' of the target C is a G, in which case editing efficiency was substantially lower (Figures 12A to 12B). NBEl activity *in vitro* was assessed on all four NC motifs at positions 1 through 8 within the protospacer (Figures 12A to 12B). In general, NBEl activity on substrates was observed to follow the order TC ≥ CC ≥ AC >GC, with maximum editing efficiency achieved when the target C is at or near position 7. In addition, it was observed that the nucleobase editor is highly processive, and will efficiently convert most of all Cs to Us on the same DNA strand within the 5-base activity window (Figure 17).

[00414] While BE1 efficiently processes substrates in a test tube, in cells a tree of possible DNA repair outcomes determines the fate of the initial U:G product of base editing (Figure 29A). To test the effectiveness of nucleobase editing in human cells, NBEl codon usage was optimized for mammalian expression, appended a C-terminal nuclear localization sequence (NLS),[53] and assayed its ability to convert C to T in human cells on 14Cs in six well-studied target sites throughout the human genome (Figure 37A).[54] The editable Cs were confirmed within each protospacer *in vitro* by incubating NBEl with synthetic 80-mers that correspond to the six different genomic sites, followed by HTS (Figures 13A to 13C, Figure 29B and Figure 25). Next, HEK293T cells were transfected with plasmids encoding NBEl and one of the six target sgRNAs, allowed three days for nucleobase editing to occur, extracted genomic DNA from the cells, and analyzed the loci by HTS. Although C to T editing in cells at the target locus was observed for all six cases, the efficiency of nucleobase editing was 1.1% to 6.3% or 0.8%-7.7% of total DNA sequences (corresponding to 2.2% to 12.6% of targeted strands), a 6.3-fold to 37-fold or 5-fold to 36-fold decrease in efficiency compared to that of

in vitro nucleobase editing (Figures 13A to 13C, Figure 29B and Figure 25). It was observed that some base editing outside of the typical window of positions 4 to 8 when the substrate C is preceded by a T, which we attribute to the unusually high activity of APOBEC1 for TC substrates.[48]

[00415] It was asked whether the cellular DNA repair response to the presence of U:G heteroduplex DNA was responsible for the large decrease in nucleobase editing efficiency in cells (Figure 29A). Uracil DNA glycosylase (UDG) catalyzes removal of U from DNA in cells and initiates base excision repair (BER), with reversion of the U:G pair to a C:G pair as the most common outcome (Figure 29A).[55] Uracil DNA glycosylase inhibitor (UGI), an 83-residue protein from *B. subtilis* bacteriophage PBS1, potently blocks human UDG activity (IC_{50} = 12 pM).[56] UGI was fused to the C-terminus of NBEl to create the second-generation nucleobase editor NBE2 and repeated editing assays on all six genomic loci. Editing efficiencies in human cells were on average 3-fold higher with NBE2 than with NBEl, resulting in gene conversion efficiencies of up to 22.8% of total DNA sequenced (up to 45.6% of targeted strands) (Figures 13A to 13C and Figure 29B). To test base editing in human cells, BE1 codon usage was optimized for mammalian expression and appended a C-terminal nuclear localization sequence (NLS).[53]

[00416] Similar editing efficiencies were observed when a separate plasmid overexpressing UGI was co-transfected with NBEl (Figures 18A to 18H). However, while the direct fusion of UGI to NBEl resulted in no significant increase in C to T mutations at monitored non-targeted genomic locations, overexpression of unfused UGI detectably increased the frequency of C to T mutations elsewhere in the genome (Figures 18A to 18H). The generality of NBE2-mediated nucleobase editing was confirmed by assessing editing efficiencies on the same six genomic targets in U20S cells, and observed similar results with those in HEK293T cells (Figure 19). Importantly, NBE2 typically did not result in any detectable indels (Figure 13C and Figure 29C), consistent with the known mechanistic dependence of NHEJ on double-stranded DNA breaks.[57, 78] Together, these results indicate that conjugating UGI to NBEl can greatly increase the efficiency of nucleobase editing in human cells.

[00417] The permanence of nucleobase editing in human cells was confirmed by monitoring editing efficiencies over multiple cell divisions in HEK293T cells at two of the tested genomic loci. Genomic DNA was harvested at two time points: three days after transfection with plasmids expressing NBE2 and appropriate sgRNAs, and after passaging the cells and growing them for four additional days (approximately five subsequent cell divisions). No significant change in editing efficiency was observed between the non-passaged cells (editing observed in

4.6% to 6.6% of targeted strands for three different target Cs) and passaged cells (editing observed in 4.6% to 6.4% of targeted strands for the same three target Cs), confirming that the nucleobase edits became permanent following cell division (Figure 20). Indels will on rare occasion arise from the processing of U:G lesions by cellular repair processes, which involve single-strand break intermediates that are known to lead to indels.[84] Given that several hundred endogenous U:G lesions are generated every day per human cell from spontaneous cytidine deaminase,[85] it was anticipate that the total indel frequency from U:G lesion repair is unlikely to increase from BE1 or BE2 activity at a single target locus.

[00418] To further increase the efficiency of nucleobase editing in cells, it was anticipated that nicking the non-edited strand may result in a smaller fraction of edited Us being removed by the cell, since eukaryotic mismatch repair machinery uses strand discontinuity to direct DNA repair to any broken strand of a mismatched duplex (Figure 29A).[58, 79, 80] The catalytic His residue was restored at position 840 in the Cas9 HNH domain,[47,59] resulting in the third-generation nucleobase editor NBE3 that nicks the non-edited strand containing a G opposite the targeted C, but does not cleave the target strand containing the C. Because NBE3 still contains the AsplOAla mutation in Cas9, it does not induce double-stranded DNA cleavage. This strategy of nicking the non-edited strand augmented nucleobase editing efficiency in human cells by an additional 1.4- to 4.8-fold relative to NBE2, resulting in up to 36.3% of total DNA sequences containing the targeted C to T conversion on the same six human genomic targets in HEK293T cells (Figures 13A to 13C and Figure 29B). Importantly, only a small frequency of indels, averaging 0.8% (ranging from 0.2% to 1.6% for the six different loci), was observed from NBE3 treatment (Figure 13C, Figure 29C, and Figure 34). In contrast, when cells were treated with wild-type Cas9, sgRNA, and a single-stranded DNA donor template to mediate HDR at three of these loci C to T conversion efficiencies averaging only 0.7% were observed, with much higher relative indel formation averaging 3.9% (Figures 13A to 13C and Figure 29C). The ratio of allele conversion to NHEJ outcomes averaged >1,000 for BE2, 23 for BE3, and 0.17 for wild-type Cas9 (Fig. 3c). We confirmed the permanence of base editing in human cells by monitoring editing efficiencies over multiple cell divisions in HEK293T cells at the HEK293 site 3 and 4 genomic loci (Figure 38). These results collectively establish that nucleobase editing can effect much more efficient targeted single-base editing in human cells than Cas9-mediated HDR, and with much less (NBE3) or no (NBE2) indel formation.

[00419] Next, the off-target activity of NBE1, NBE2, and NBE3 in human cells was evaluated. The off-target activities of Cas9, dCas9, and Cas9 nickase have been extensively

studied (Figures 23 to 24 and 31 to 33).[54,60-62] Because the sequence preference of rAPOBECl has been shown to be independent of DNA bases more than one base from the target C,[63] consistent with the sequence context independence observed in Figures 12A to 12B, it was assumed that potential off-target activity of nucleobase editors arises from off-target Cas9 binding. Since only a fraction of Cas9 off-target sites will have a C within the active window for nucleobase editing, off-target nucleobase editing sites should be a subset of the off-target sites of canonical Cas9 variants. For each of the six sites studied, the top ten known Cas9 off-target loci in human cells that were previously determined using the GUIDE-seq method were sequenced (Figures 23 to 27 and 31 to 33).[54,61] Detectable off-target nucleobase editing at only a subset (16/34, 47% for NBEl and NBE2, and 17/34, 50% for NBE3) of known dCas9 off-target loci was observed. In all cases, the off-target base-editing substrates contained a C within the five-base target window. In general, off-target C to T conversion paralleled off-target Cas9 nuclease-mediated genome modification frequencies (Figures 23 to 27). Also monitored were C to T conversions at 2,500 distinct cytosines surrounding the six on-target and 34 off-target loci tested, representing a total of 14,700,000 sequence reads derived from approximately 1.8xl0 ^{6} cells, and observed no detectable increase in C to T conversions at any of these other sites upon NBEl, NBE2, or NBE3 treatment compared to that of untreated cells (Figure 28). Taken together, these findings suggest that off-target substrates of nucleobase editors include a subset of Cas9 off-target substrates, and that nucleobase editors in human cells do not induce untargeted C to T conversion throughout the genome at levels that can be detected by the methods used here. No substantial change was observed in editing efficiency between non-passaged HEK293T cells (editing observed in 1.8% to 2.6% of sequenced strands for the three target Cs with BE2, and 6.2% to 14.3% with BE3) and cells that had undergone approximately five cell divisions after base editing (editing observed in 1.9% to 2.3% of sequenced strands for the same target Cs with BE2, and 6.4% to 14.5% with BE3), confirming that base edits in these cells are durable (Extended Data Fig. 6).

[00420] Finally, the potential of nucleobase editing to correct three disease-relevant mutations in mammalian cells was tested. The apolipoprotein E gene variant *APOE4* encodes two Arg residues at amino acid positions 112 and 158, and is the largest and most common genetic risk factor for late-onset Alzheimer's disease.[64] ApoE variants with Cys residues in positions 112 or 158, including *APOE2* (Cysll2/Cysl58), *APOE3* (Cysll2/Argl58), and *APOE3'* (Argll2/Cysl58) have been shown[65] or are presumed[81] to confer substantially lower Alzheimer's disease risk than *APOE4.* Encouraged by the ability of NBEl to convert *APOE4* to *APOE3' in vitro* (Figures 16A to 16B), this conversion was attempted in immortalized

mouse astrocytes in which the endogenous murine *APOE* gene has been replaced by human *APOE4* (Taconic). DNA encoding NBE3 and an appropriate sgRNA was delivered into these astrocytes by nucleofection (nucleofection efficiency of 25%), extracted genomic DNA from all treated cells two days later, and measured editing efficiency by HTS. Conversion of Argl58 to Cysl58 was observed in 58-75% of total DNA sequencing reads (44% of nucleofected astrocytes) (Figures 14A to 14C and Figures 30A). Also observed was 36-50% editing of total DNA at the third position of codon 158 and 38-55% editing of total DNA at the first position of Leu 159, as expected since all three of these Cs are within the active nucleobase editing window. However, neither of the other two C→T conversions results in a change in the amino acid sequence of the ApoE3' protein since both TGC and TGT encode Cys, and both CTG and TTG encode Leu. From > 1,500,000 sequencing reads derived from lxlO^6 cells evidence of 1.7% indels at the targeted locus following NBE3 treatment was observed (Figure 35). In contrast, identical treatment of astrocytes with wt Cas9 and donor ssDNA resulted in 0.1-0.3% APOE4 correction and 26-40% indels at the targeted locus, efficiencies consistent with previous reports of single-base correction using Cas9 and HDR[45,75] (Figure 30A and Figure 40A). Astrocytes treated identically but with an sgRNA targeting the VEGFA locus displayed no evidence of APOE4 base editing (Figure 34 and Figure 40A). These results demonstrate how nucleobase editors can effect precise, single-amino acid changes in the coding sequence of a protein as the major product of editing, even when their processivity results in more than one nucleotide change in genomic DNA. The off-target activities of Cas9, dCas9, and Cas9 nickase have been extensively studied.[54,60-62] In general, off-target C to T conversions by BE1, BE2, and BE3 paralleled off-target Cas9 nuclease-mediated genome modification frequencies.

[00421] The dominant-negative p53 mutations Tyrl63Cys and Asn239Asp are strongly associated with several types of cancer.[66-67] Both of these mutations can be corrected by a C to T conversion on the template strand (Figures 16A to 16B). A human breast cancer cell line homozygous for the p53 Tyrl63Cys mutation (HCC1954 cells) was nucleofected with DNA encoding NBE3 and an sgRNA programmed to correct Tyrl63Cys. Because the nucleofection efficiency of HCC1954 cells was < 10%, a plasmid expressing IRFP was co-nucleofected into these cells to enable isolation of nucleofected cells by fluorescence-activated cell sorting two days after treatment. HTS of genomic DNA revealed correction of the Tyrl63Cys mutation in 7.6% of nucleofected HCC1954 cells (Figure 30B and Figure 40A to 40B). Also nucleofected was a human lymphoma cell line that is heterozygous for p53 Asn239Asp (ST486 cells) with DNA encoding NBE2 and an sgRNA programmed to correct Asn239Asp with 92%

nucleofection efficiency). Correction of the Asn239Asp mutation was observed in 11% of treated ST486 cells (12% of nucleofected ST486 cells). Consistent with the findings in HEK cells, no indels were observed from the treatment of ST486 cells with NBE2, and 0.6% indel formation from the treatment of HCC1954 cells with NBE3. No other DNA changes within at least 50 base pairs of both sides of the protospacer were detected at frequencies above that of untreated controls out of > 2,000,000 sequencing reads derived from 2×10^5 cells (Figures 14A to 14C, Figure 30B). These results collectively represent the conversion of three disease-associated alleles in genomic DNA into their wild-type forms with an efficiency and lack of other genome modification events that is, to our knowledge, not currently achievable using other methods.

[00422] To illuminate the potential relevance of nucleobase editors to address human genetic diseases, the NCBI ClinVar database[68] was searched for known genetic diseases that could in principle be corrected by this approach. ClinVar was filtered by first examining only single nucleotide polymorphisms (SNPs), then removing any nonpathogenic variants. Out of the 24,670 pathogenic SNPs, 3,956 are caused by either a T to C, or an A to G, substitution. This list was further filtered to only include variants with a nearby NGG PAM that would position the SNP within the deamination activity window, resulting in 1,089 clinically relevant pathogenic gene variants that could in principle be corrected by the nucleobase editors described here (Figure 21).

[00423] In some embodiments, any of the base editors provided herein may be used to treat a disease or disorder. For example, any base editors provided herein may be used to correct one or more mutations associated with any of the diseases or disorders provided herein. Exemplary diseases or disorders that may be treated include, without limitation, 3-Methylglutaconic aciduria type 2, 46,XY gonadal dysgenesis, 4-Alpha-hydroxyphenylpyruvate hydroxylase deficiency, 6-pyruvoyl-tetrahydropterin synthase deficiency, achromatopsia, Acid-labile subunit deficiency, Acrodysostosis, acroerythrokeratoderma, ACTH resistance, ACTH-independent macronodular adrenal hyperplasia, Activated PI3K-delta syndrome, Acute intermittent porphyria, Acute myeloid leukemia, Adams-Oliver syndrome 1/5/6, Adenylosuccinate lyase deficiency, Adrenoleukodystrophy, Adult neuronal ceroid lipofuscinosis, Adult onset ataxia with oculomotor apraxia, Advanced sleep phase syndrome, Age-related macular degeneration, Alagille syndrome, Alexander disease, Allan-Herndon-Dudley syndrome, Alport syndrome, X-linked recessive, Alternating hemiplegia of childhood, Alveolar capillary dysplasia with misalignment of pulmonary veins, Amelogenesis imperfecta, Amyloidogenic transthyretin

amyloidosis, Amyotrophic lateral sclerosis, Anemia (nonspherocytic hemolytic, due to G6PD deficiency), Anemia (sideroblastic, pyridoxine-refractory, autosomal recessive), Anonychia, Antithrombin III deficiency, Aortic aneurysm, Aplastic anemia, Apolipoprotein C2 deficiency, Apparent mineralocorticoid excess, Aromatase deficiency, Arrhythmogenic right ventricular cardiomyopathy, Familial hypertrophic cardiomyopathy, Hypertrophic cardiomyopathy, Arthrogryposis multiplex congenital, Aspartylglycosaminuria, Asphyxiating thoracic dystrophy, Ataxia with vitamin E deficiency, Ataxia (spastic), Atrial fibrillation, Atrial septal defect, atypical hemolytic-uremic syndrome, autosomal dominant CD11C+/CD1C+ dendritic cell deficiency, Autosomal dominant progressive external ophthalmoplegia with mitochondrial DNA deletions, Baraitser-Winter syndrome, Bartter syndrome, Basa ganglia calcification, Beckwith-Wiedemann syndrome, Benign familial neonatal seizures, Benign scapuloperoneal muscular dystrophy, Bernard Soulier syndrome, Beta thalassemia intermedia, Beta-D-mannosidosis, Bietti crystalline corneoretinal dystrophy, Bile acid malabsorption, Biotinidase deficiency, Borjeson-Forssman-Lehmann syndrome, Boucher Neuhauser syndrome, Bowen-Conradi syndrome, Brachydactyly, Brown-Vialetto-Van laere syndrome, Brugada syndrome, Cardiac arrhythmia, Cardiofaciocutaneous syndrome, Cardiomyopathy, Carnevale syndrome, Carnitine palmitoyltransferase II deficiency, Carpenter syndrome, Cataract, Catecholaminergic polymorphic ventricular tachycardia, Central core disease, Centromeric instability of chromosomes 1,9 and 16 and immunodeficiency, Cerebral autosomal dominant arteriopathy, Cerebro-oculo-facio-skeletal syndrome, Ceroid lipofuscinosis, Charcot-Marie-Tooth disease, Cholestanol storage disease, Chondrocalcinosis, Chondrodysplasia, Chronic progressive multiple sclerosis, Coenzyme Q10 deficiency, Cohen syndrome, Combined deficiency of factor V and factor VIII, Combined immunodeficiency, Combined oxidative phosphorylation deficiency, Combined partial 17-alpha-hydroxylase/17,20- lyase deficiency, Complement factor d deficiency, Complete combined 17-alpha- hydroxylase/ 17,20-lyase deficiency, Cone-rod dystrophy, Congenital contractural arachnodactyly, Congenital disorder of glycosylation, Congenital lipomatous overgrowth, Neoplasm of ovary, PIK3CA Related Overgrowth Spectrum, Congenital long QT syndrome, Congenital muscular dystrophy, Congenital muscular hypertrophy-cerebral syndrome, Congenital myasthenic syndrome, Congenital myopathy with fiber type disproportion, Eichsfeld type congenital muscular dystrophy, Congenital stationary night blindness, Corneal dystrophy, Cornelia de Lange syndrome, Craniometaphyseal dysplasia, Crigler Najjar syndrome, Crouzon syndrome, Cutis laxa with osteodystrophy, Cyanosis, Cystic fibrosis, Cystinosis, Cytochrome-c oxidase deficiency, Mitochondrial complex I deficiency, D-2-hydroxyglutaric aciduria, Danon disease, Deafness

with labyrinthine aplasia microtia and microdontia (LAMM), Deafness, Deficiency of acetyl-CoA acetyltransferase, Deficiency of ferroxidase, Deficiency of UDPglucose-hexose-1-phosphate uridylyltransferase, Dejerine-Sottas disease, Desbuquois syndrome, DFNA, Diabetes mellitus type 2, Diabetes-deafness syndrome, Diamond-Blackfan anemia, Diastrophic dysplasia, Dihydropteridine reductase deficiency, Dihydropyrimidinase deficiency, Dilated cardiomyopathy, Disseminated atypical mycobacterial infection, Distal arthrogryposis, Distal hereditary motor neuronopathy, Donnai Barrow syndrome, Duchenne muscular dystrophy, Becker muscular dystrophy, Dyschromatosis universalis hereditaria, Dyskeratosis congenital, Dystonia, Early infantile epileptic encephalopathy, Ehlers-Danlos syndrome, Eichsfeld type congenital muscular dystrophy, Emery-Dreifuss muscular dystrophy, Enamel-renal syndrome, Epidermolysis bullosa dystrophica inversa, Epidermolysis bullosa herpetiformis, Epilepsy, Episodic ataxia, Erythrokeratodermia variabilis, Erythropoietic protoporphyria, Exercise intolerance, Exudative vitreoretinopathy, Fabry disease, Factor V deficiency, Factor VII deficiency, Factor xiii deficiency, Familial adenomatous polyposis, breast cancer, ovarian cancer, cold urticarial, chronic infantile neurological, cutaneous and articular syndrome, hemiplegic migraine, hypercholesterolemia, hypertrophic cardiomyopathy, hypoalphalipoproteinemia, hypokalemia-hypomagnesemia, juvenile gout, hyperlipoproteinemia, visceral amyloidosis, hypophosphatemic vitamin D refractory rickets, FG syndrome, Fibrosis of extraocular muscles, Finnish congenital nephrotic syndrome, focal epilepsy, Focal segmental glomerulosclerosis, Frontonasal dysplasia, Frontotemporal dementia, Fructose-biphosphatase deficiency, Gamstorp-Wohlfart syndrome, Ganglioside sialidase deficiency, GATA-1-related thrombocytopenia, Gaucher disease, Giant axonal neuropathy, Glanzmann thrombasthenia, Glomerulocystic kidney disease, Glomerulopathy, Glucocorticoid resistance, Glucose-6-phosphate transport defect, Glutaric aciduria, Glycogen storage disease, Gorlin syndrome, Holoprosencephaly, GRACILE syndrome, Haemorrhagic telangiectasia, Hemochromatosis, Hemoglobin H disease, Hemolytic anemia, Hemophagocytic lymphohistiocytosis, Carcinoma of colon, Myhre syndrome, leukoencephalopathy, Hereditary factor IX deficiency disease, Hereditary factor VIII deficiency disease, Hereditary factor XI deficiency disease, Hereditary fructosuria, Hereditary Nonpolyposis Colorectal Neoplasm, Hereditary pancreatitis, Hereditary pyropoikilocytosis, Elliptocytosis, Heterotaxy, Heterotopia, Histiocytic medullary reticulosis, Histiocytosis-lymphadenopathy plus syndrome, HNSHA due to aldolase A deficiency, Holocarboxylase synthetase deficiency, Homocysteinemia, Howel-Evans syndrome, Hydatidiform mole, Hypercalciuric hypercalcemia, Hyperimmunoglobulin D, Mevalonic aciduria,

Hyperinsulinemic hypoglycemia, Hyperkalemic Periodic Paralysis, Paramyotonia congenita of von Eulenburg, Hyperlipoproteinemia, Hypermanganesemia, Hypermethioninemia, Hyperphosphatasemia, Hypertension, hypomagnesemia, Hypobetalipoproteinemia, Hypocalcemia, Hypogonadotropic hypogonadism, Hypogonadotropic hypogonadism, Hypohidrotic ectodermal dysplasia, Hyper-IgM immunodeficiency, Hypohidrotic X-linked ectodermal dysplasia, Hypomagnesemia, Hypoparathyroidism, Idiopathic fibrosing alveolitis, Immunodeficiency, Immunoglobulin A deficiency, Infantile hypophosphatasia, Infantile Parkinsonism-dystonia, Insulin-dependent diabetes mellitus, Intermediate maple syrup urine disease, Ischiopatellar dysplasia, Islet cell hyperplasia, Isolated growth hormone deficiency, Isolated lutropin deficiency, Isovaleric acidemia, Joubert syndrome, Juvenile polyposis syndrome, Juvenile retinoschisis, Kallmann syndrome, Kartagener syndrome, Kugelberg-Welander disease, Lattice corneal dystrophy, Leber congenital amaurosis, Leber optic atrophy, Left ventricular noncompaction, Leigh disease, Mitochondrial complex I deficiency, Leprechaunism syndrome, Arthrogryposis, Anterior horn cell disease, Leukocyte adhesion deficiency, Leukodystrophy, Leukoencephalopathy, Ovarioleukodystrophy, L-ferritin deficiency, Li-Fraumeni syndrome, Limb-girdle muscular dystrophy- dystroglycanopathy, Loeys-Dietz syndrome, Long QT syndrome, Macrocephaly/autism syndrome, Macular corneal dystrophy, Macular dystrophy, Malignant hyperthermia susceptibility, Malignant tumor of prostate, Maple syrup urine disease, Marden Walker like syndrome, Marfan syndrome, Marie Unna hereditary hypotrichosis, Mast cell disease, Meconium ileus, Medium-chain acyl-coenzyme A dehydrogenase deficiency, Melnick-Fraser syndrome, Mental retardation, Merosin deficient congenital muscular dystrophy, Mesothelioma, Metachromatic leukodystrophy, Metaphyseal chondrodysplasia, Methemoglobinemia, methylmalonic aciduria, homocystinuria, Microcephaly, chorioretinopathy, lymphedema, Microphthalmia, Mild non-PKU hyperphenylalanemia, Mitchell-Riley syndrome, mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase deficiency, Mitochondrial complex I deficiency, Mitochondrial complex III deficiency, Mitochondrial myopathy, Mucolipidosis III, Mucopolysaccharidosis, Multiple sulfatase deficiency, Myasthenic syndrome, Mycobacterium tuberculosis, Myeloperoxidase deficiency, Myhre syndrome, Myoclonic epilepsy, Myofibrillar myopathy, Myoglobinuria, Myopathy, Myopia, Myotonia congenital, Navajo neurohepatopathy, Nemaline myopathy, Neoplasm of stomach, Nephrogenic diabetes insipidus, Nephronophthisis, Nephrotic syndrome, Neurofibromatosis, Neutral lipid storage disease, Niemann-Pick disease, Non-ketotic hyperglycinemia, Noonan syndrome, Noonan syndrome-like disorder, Norum disease, Macular degeneration, N-terminal acetyltransferase deficiency,

Oculocutaneous albinism, Oculodentodigital dysplasia, Ohdo syndrome, Optic nerve aplasia, Ornithine carbamoyltransferase deficiency, Orofaciodigital syndrome, Osteogenesis imperfecta, Osteopetrosis, Ovarian dysgenesis, Pachyonychia, Palmoplantar keratoderma, nonepidermolytic, Papillon-Lef\xc3\xa8vre syndrome, Haim-Munk syndrome, Periodontitis, Peeling skin syndrome, Pendred syndrome, Peroxisomal fatty acyl-coa reductase 1 disorder, Peroxisome biogenesis disorder, Pfeiffer syndrome, Phenylketonuria, Phenylketonuria, Hyperphenylalaninemia, non- PKU, Pituitary hormone deficiency, Pityriasis rubra pilaris, Polyarteritis nodosa, Polycystic kidney disease, Polycystic lipomembranous osteodysplasia, Polymicrogyria, Pontocerebellar hypoplasia, Porokeratosis, Posterior column ataxia, Primary erythromelalgia, hyperoxaluria, Progressive familial intrahepatic cholestasis, Progressive pseudorheumatoid dysplasia, Propionic acidemia, Pseudohermaphroditism, Pseudohypoaldosteronism, Pseudoxanthoma elasticum-like disorder, Purine-nucleoside phosphorylase deficiency, Pyridoxal 5-phosphate-dependent epilepsy, Renal dysplasia, retinal pigmentary dystrophy, cerebellar ataxia, skeletal dysplasia, Reticular dysgenesis, Retinitis pigmentosa, Usher syndrome, Retinoblastoma, Retinopathy, RRM2B-related mitochondrial disease, Rubinstein-Taybi syndrome, Schnyder crystalline corneal dystrophy, Sebaceous tumor, Severe congenital neutropenia, Severe myoclonic epilepsy in infancy, Severe X-linked myotubular myopathy, onychodysplasia, facial dysmorphism, hypotrichosis, Short-rib thoracic dysplasia, Sialic acid storage disease, Sialidosis, Sideroblastic anemia, Small fiber neuropathy, Smith-Magenis syndrome, Sorsby fundus dystrophy, Spastic ataxia, Spastic paraplegia, Spermatogenic failure, Spherocytosis, Sphingomyelin/cholesterol lipidosis, Spinocerebellar ataxia, Split-hand/foot malformation, Spondyloepimetaphyseal dysplasia, Platyspondylic lethal skeletal dysplasia, Squamous cell carcinoma of the head and neck, Stargardt disease, Sucrase-isomaltase deficiency, Sudden infant death syndrome, Supravalvar aortic stenosis, Surfactant metabolism dysfunction, Tangier disease, Tatton-Brown-rahman syndrome, Thoracic aortic aneurysms and aortic dissections, Thrombophilia, Thyroid hormone resistance, TNF receptor-associated periodic fever syndrome (TRAPS), Tooth agenesis, Torsades de pointes, Transposition of great arteries, Treacher Collins syndrome, Tuberous sclerosis syndrome, Tyrosinase-negative oculocutaneous albinism, Tyrosinase-positive oculocutaneous albinism, Tyrosinemia, UDPglucose-4-epimerase deficiency, Ullrich congenital muscular dystrophy, Bethlem myopathy Usher syndrome, UV-sensitive syndrome, Van der Woude syndrome, popliteal pterygium syndrome, Very long chain acyl-CoA dehydrogenase deficiency, Vesicoureteral reflux, Vitreoretinochoroidopathy, Von Hippel-Lindau syndrome, von Willebrand disease, Waardenburg syndrome, Warsaw breakage syndrome, WFSl -Related

Disorders, Wilson disease, Xeroderma pigmentosum, X-linked agammaglobulinemia, X-linked hereditary motor and sensory neuropathy, X-linked severe combined immunodeficiency, and Zellweger syndrome.

[00424] The development of nucleobase editing advances both the scope and effectiveness of genome editing. The nucleobase editors described here offer researchers a choice of editing with virtually no indel formation (NBE2), or more efficient editing with a low frequency (here, typically $\leq$ 1%) of indel formation (NBE3). That the product of base editing is, by definition, no longer a substrate likely contributes to editing efficiency by preventing subsequent product transformation, which can hamper traditional Cas9 applications. By removing the reliance on double-stranded DNA cleavage and stochastic DNA repair processes that vary greatly by cell state and cell type, nucleobase editing has the potential to expand the type of genome modifications that can be cleanly installed, the efficiency of these modifications, and the type of cells that are amenable to editing. It is likely that recent engineered Cas9 variants[69,70, 82] or delivery methods[71] with improved DNA specificity, as well as Cas9 variants with altered PAM specificities,[72] can be integrated into this strategy to provide additional nucleobase editors with improved DNA specificity or that can target an even wider range of disease-associated mutations. These findings also suggest that engineering additional fusions of dCas9 with enzymes that catalyze additional nucleobase transformations will increase the fraction of the possible DNA base changes that can be made through nucleobase editing. These results also suggest architectures for the fusion of other DNA-modifying enzymes, including methylases and demathylases, that mau enable additional types of programmable genome and epigenome base editing.

Materials and Methods

[00425] **Cloning.** DNA sequences of all constructs and primers used in this paper are listed in the Supplementary Sequences. Plasmids containing genes encoding NBE1, NBE2, and NBE3 will be available from Addgene. PCR was performed using VeraSeq ULtra DNA polymerase (Enzymatics), or Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs). NBE plasmids were constructed using USER cloning (New England Biolabs). Deaminase genes were synthesized as gBlocks Gene Fragments (Integrated DNA Technologies), and Cas9 genes were obtained from previously reported plasmids.[18] Deaminase and fusion genes were cloned into pCMV (mammalian codon-optimized) or pET28b (*E. coli* codon-optimized) backbones. sgRNA expression plasmids were constructed using site-

directed mutagenesis. Briefly, the primers listed in the Supplementary Sequences were 5' phosphorylated using T4 Polynucleotide Kinase (New England Biolabs) according to the manufacturer's instructions. Next, PCR was performed using Q5 Hot Start High-Fidelity Polymerase (New England Biolabs) with the phosphorylated primers and the plasmid pFYF1320 (EGFP sgRNA expression plasmid) as a template according to the manufacturer's instructions. PCR products were incubated with Dpnl (20 U, New England Biolabs) at 37 °C for 1 h, purified on a QIAprep spin column (Qiagen), and ligated using QuickLigase (New England Biolabs) according to the manufacturer's instructions. DNA vector amplification was carried out using Machl competent cells (ThermoFisher Scientific).

[00426] ***In vitro*** **deaminase assay on ssDNA.** Sequences of all ssDNA substrates are listed in the Supplementary Sequences. All Cy3-labelled substrates were obtained from Integrated DNA Technologies (IDT). Deaminases were expressed *in vitro* using the TNT T7 Quick Coupled Transcription/Translation Kit (Promega) according to the manufacturer' s instructions using 1 μg of plasmid. Following protein expression, 5 μL of lysate was combined with 35 μL of ssDNA (1.8 μM) and USER enzyme (1 unit) in CutSmart buffer (New England Biolabs) (50 mM potassium acetate, 29 mM Trisacetate, 10 mM magnesium acetate, 100 ug/mL BSA, pH 7.9) and incubated at 37 °C for 2 h. Cleaved U-containing substrates were resolved from full-length unmodified substrates on a 10% TBE-urea gel (Bio-Rad).

[00427] **Expression and purification of His_6-rAPOBECl-linker-dCas9 fusions.** *E. Coli* BL21 STAR (DE3)-competent cells (ThermoFisher Scientific) were transformed with plasmids encoding pET28b-His_6-rAPOBEC-linker-dCas9 with GGS, $(GGS)_3$, (SEQ ID NO: 610) XTEN, or $(GGS)_7$ (SEQ ID NO: 610) linkers. The resulting expression strains were grown overnight in Luria-Bertani (LB) broth containing 100 μg/mL of kanamycin at 37 °C. The cells were diluted 1:100 into the same growth medium and grown at 37 °C to OD_{600} = -0.6. The culture was cooled to 4 °C over a period of 2 h, and isopropyl -β-D-l-thiogalactopyranoside (IPTG) was added at 0.5 mM to induce protein expression. After -16 h, the cells were collected by centrifugation at 4,000 *g* and resuspended in lysis buffer (50 mM tris(hydroxymethyl)-amiiiomethane (Tris)-HCl, pH 7.0, 1 M NaCi, 20% glycerol, 10 mM tris(2-carboxyethyl)phosphine (TCEP, Soltec Ventures)). The cells were lysed by sonication (20 s pulse-on, 20 s pulse-off for 8 min total at 6 W output) and the lysate supernatant was isolated following centrifugation at 25,000 *g* for 15 min. The lysate was incubated with His-Pur nickel-nitriloacetic acid (nickel-NTA) resin (ThermoFisher Scientific) at 4 °C for 1 h to capture the His-tagged fusion protein. The resin was transferred to a column and washed with 40 mL of lysis buffer. The His-tagged fusion protein was eluted in lysis buffer supplemented

with 285 mM imidazole, and concentrated by ultrafiltration (Amicon-Millipore, 100-kDa molecular weight cut-off) to 1 mL total volume. The protein was diluted to 20 mL in low-salt purification buffer containing 50 mM tris(hydroxymeiliyl)-aminomeihane (Tris)-HCl, pH 7.0, 0.1 M NaCl, 20% glycerol, 10 mM TCEP and loaded onto SP Sepharose Fast Flow resin (GE Life Sciences). The resin was washed with 40 mL of this low-salt buffer, and the protein eluted with 5 mL of activity buffer containing 50 mM tris(hydroxymethyl)-aminomethaxe (Tris)-HCl, pH 7.0, 0.5 M NaCl, 20% glycerol, 10 mM TCEP. The eluted proteins were quantified on a SDSPAGE gel.

[00428] *In vitro* transcription of sgRNAs. Linear DNA fragments containing the T7 promoter followed by the 20-bp sgRNA target sequence were transcribed *in vitro* using the primers listed in the Supplementary Sequences with the TranscriptAid T7 High Yield Transcription Kit (ThermoFisher Scientific) according to the manufacturer's instructions. sgRNA products were purified using the MEGAclear Kit (ThermoFisher Scientific) according to the manufacturer's instructions and quantified by UV absorbance.

[00429] **Preparation of Cy3-conjugated dsDNA substrates.** Sequences of 80-nucleotide unlabeled strands are listed in the Supplementary Sequences and were ordered as PAGE-purified oligonucleotides from IDT. The 25-nt Cy3-labeled primer listed in the Supplementary Sequences is complementary to the 3' end of each 80-nt substrate. This primer was ordered as an HPLC-purified oligonucleotide from IDT. To generate the Cy3-labeled dsDNA substrates, the 80-nt strands (5 pL of a 100 μM solution) were combined with the Cy3-labeled primer (5 μL of a 100 pM solution) in NEBuffer 2 (38.25 pL of a 50 mM NaCl, 10 mMTris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9 solution, New England Biolabs) with dNTPs (0.75 pL of a 100 mM solution) and heated to 95 °C for 5 min, followed by a gradual cooling to 45 °C at a rate of 0.1 °C/s. After this annealing period, Klenow exo$^-$ (5 U, New England Biolabs) was added and the reaction was incubated at 37 °C for 1 h. The solution was diluted with Buffer PB (250 pL, Qiagen) and isopropanol (50 pL) and purified on a QIAprep spin column (Qiagen), eluting with 50 pL of Tris buffer.

[00430] **Deaminase assay on dsDNA.** The purified fusion protein (20 pL of 1.9 pM in activity buffer) was combined with 1 equivalent of appropriate sgRNA and incubated at ambient temperature for 5 min. The Cy3-labeled dsDNA substrate was added to final concentration of 125 nM and the resulting solution was incubated at 37 °C for 2 h. The dsDNA was separated from the fusion by the addition of Buffer PB (100 pL, Qiagen) and isopropanol (25 pL) and purified on a EconoSpin micro spin column (Epoch Life Science), eluting with 20 pL of CutSmart buffer (New England Biolabs). USER enzyme (1 U, New

England Biolabs) was added to the purified, edited dsDNA and incubated at 37 °C for 1 h. The Cy3-labeled strand was fully denatured from its complement by combining 5 μL of the reaction solution with 15 μL of a DMSO-based loading buffer (5 mM Tris, 0.5 mM EDTA, 12.5% glycerol, 0.02% bromophenol blue, 0.02% xylene cyan, 80% DMSO). The full-length C-containing substrate was separated from any cleaved, U-containing edited substrates on a 10% TBE-urea gel (Bio-Rad) and imaged on a GE Amersham Typhoon imager.

[00431] Preparation of *in vitro*-edited dsDNA for high-throughput sequencing (HTS). The oligonucleotides listed in the Supplementary Sequences were obtained from IDT. Complementary sequences were combined (5 μL of a 100 μM solution) in Tris buffer and annealed by heating to 95 °C for 5 min, followed by a gradual cooling to 45 °C at a rate of 0.1 °C/s to generate 60-bp dsDNA substrates. Purified fusion protein (20 μL of 1.9 μM in activity buffer) was combined with 1 equivalent of appropriate sgRNA and incubated at ambient temperature for 5 min. The 60-mer dsDNA substrate was added to final concentration of 125 nM and the resulting solution was incubated at 37 °C for 2 h. The dsDNA was separated from the fusion by the addition of Buffer PB (100 μL, Qiagen) and isopropanol (25 μL) and purified on a EconoSpin micro spin column (Epoch Life Science), eluting with 20 μL of Tris buffer. The resulting edited DNA (1 μL was used as a template) was amplified by PCR using the HTS primer pairs specified in the Supplementary Sequences and VeraSeq Ultra (Enzymatics) according to the manufacturer's instructions with 13 cycles of amplification. PCR reaction products were purified using RapidTips (Diffinity Genomics), and the purified DNA was amplified by PCR with primers containing sequencing adapters, purified, and sequenced on a MiSeq high-throughput DNA sequencer (Illumina) as previously described.[73]

[00432] Cell culture. HEK293T (ATCC CRL-3216), U20S (ATCC-HTB-96) and ST486 cells (ATCC) were maintained in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher) supplemented with 10% (v/v) fetal bovine serum (FBS) and penicillin/streptomycin (lx, Amresco), at 37 °C with 5% $C0_2$. HCC1954 cells (ATCC CRL-2338) were maintained in RPMI-1640 medium (ThennoFisher Scientific) supplemented as described above. Immortalized rat astrocytes containing the ApoE4 isoform of the ***APOE*** gene (Taconic Biosciences) were cultured in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThennoFisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS) and 200 μg/mL Geneticin (ThermoFisher Scientific).

[00433] Transfections. HEK293T cells were seeded on 48-well collagen-coated BioCoat plates (Corning) and transfected at approximately 85% confluency. Briefly, 750 ng of NBE and 250 ng of sgRNA expression plasmids were transfected using 1.5 μL of Lipofectamine

2000 (ThermoFisher Scientific) per well according to the manufacturer's protocol. Astrocytes, U20S, HCC1954, HEK293T and ST486 cells were transfected using appropriate AMAXA NUCLEOFECTOR™ II programs according to manufacturer's instructions. 40 ng of infrared RFP (Addgene plasmid 45457)[74] was added to the nucleofection solution to assess nucleofection efficiencies in these cell lines. For astrocytes, U20S, and ST486 cells, nucleofection efficiencies were 25%, 74%, and 92%, respectively. For HCC1954 cells, nucleofection efficiency was <10%. Therefore, following trypsinization, the HCC1954 cells were filtered through a 40 micron strainer (Fisher Scientific), and the nucleofected HCC1954 cells were collected on a Beckman Coulter MoFlo XDP Cell Sorter using the iRFP signal (abs 643 nm, em 670 nm). The other cells were used without enrichment of nucleofected cells.

[00434] High-throughput DNA sequencing of genomic DNA samples. Transfected cells were harvested after 3 d and the genomic DNA was isolated using the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) according to the manufacturer's instructions. On-target and off-target genomic regions of interest were amplified by PCR with flanking HTS primer pairs listed in the Supplementary Sequences. PCR amplification was carried out with Phusion high-fidelity DNA polymerase (ThermoFisher) according to the manufacturer's instructions using 5 ng of genomic DNA as a template. Cycle numbers were determined separately for each primer pair as to ensure the reaction was stopped in the linear range of amplification (30, 28, 28, 28, 32, and 32 cycles for EMXl, FANCF, HEK293 site 2, HEK293 site 3, HEK293 site 4, and RNF2 primers, respectively). PCR products were purified using RapidTips (Diffinity Genomics). Purified DNA was amplified by PCR with primers containing sequencing adaptors. The products were gel-purified and quantified using the QUANT-IT™ PicoGreen dsDNA Assay Kit (ThermoFisher) and KAPA Library Quantification Kit-Illumina (KAPA Biosystems). Samples were sequenced on an Illumina MiSeq as previously described.[73]

[00435] Data analysis. Sequencing reads were automatically demultiplexed using MiSeq Reporter (Illumina), and individual FASTQ files were analyzed with a custom Matlab script provided in the Supplementary Notes. Each read was pairwise aligned to the appropriate reference sequence using the Smith-Waterman algorithm. Base calls with a Q-score below 31 were replaced with N's and were thus excluded in calculating nucleotide frequencies. This treatment yields an expected MiSeq base-calling error rate of approximately 1 in 1,000. Aligned sequences in which the read and reference sequence contained no gaps were stored in an alignment table from which base frequencies could be tabulated for each locus.

[00436] Indel frequencies were quantified with a custom Matlab script shown in the Supplementary Notes using previously described criteria[71]. Sequencing reads were scanned for exact matches to two 10-bp sequences that flank both sides of a window in which indels might occur. If no exact matches were located, the read was excluded from analysis. If the length of this indel window exactly matched the reference sequence the read was classified as not containing an indel. If the indel window was two or more bases longer or shorter than the reference sequence, then the sequencing read was classified as an insertion or deletion, respectively.

[00437] All publications, patents, patent applications, publication, and database entries *(e.g.,* sequence database entries) mentioned herein, *e.g.,* in the Background, Summary, Detailed Description, Examples, and/or References sections, are hereby incorporated by reference in their entirety as if each individual publication, patent, patent application, publication, and database entry was specifically and individually incorporated herein by reference. In case of conflict, the present application, including any definitions herein, will control.

Supplementary Sequences

[00438] **Primers used for generating sgRNA transfection plasmids.** rev_sgRNA_plasmid was used in all cases. The pFYF1320 plasmid was used as template as noted in Materials and Methods section. SEQ ID NOs: 187-196 appear from top to bottom below, respectively.

rev__sgRNA_plasmsd	GGTGTTTCGTCCTTTCCACAAG
fwd _p53 _Y163C	GCTTGCAGATGGCCATGGCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd._p53_N23QD	TGTCACACATGTAGTTGTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd._APOE4 _C158R	GAAGCGCCTGGCAGTGTACCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd._EMX1	GAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd._FANCF	GGAATCCCTTCTGCAGCACCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd._HEK293_2	GAACACAAAGCATAGACTGCG TTTTAGAG CTAGAAATAGCAAGTTAAAATAAG GG
fwd._HEK293_3	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd._HEK293_4	GGCACTGCGGCTGGAGGTGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_RNF2	GTCATCTTAGTC ATTAC CTGGTTTTAG AGCTAG AAATAGCAAGTTAAAATAAGGC

[00439] **Sequences of all ssDNA substrates used in in vitro deaminase assays.** SEQ ID NOs: 197-199 appear from top to bottom below, respectively.

rAPOBEC 1 substrate	Cy3-ATTATTATTATTCCGCGGATTT ATTTATTTATTTATTTATTT
hAiD/pmCDA 1 substrate	Cy3-ATTATTATTATTAGCTATTTATTTATTTATTTATTTATTT
hAPGBECSG substrate	Cy3-ATTATTATTATTCCCGGATTTATTTATTTATTTATTTATTT

[00440] Primers used for generating PCR products to serve as substrates for T7 transcription of sgRNAs for gel-based deaminase assay. rev_gRNA_T7 was used in all cases. The pFYF1320 plasmid was used as template as noted in Materials and Methods section. SEQ ID NOs: 200-223 appear from top to bottom below, respectively.

rev_sgRNA_T7	AAAAAAAGCACCGACTCGGTG
fwd_sgRNA._T7_dsDNA_2	TAATACGACTCACTATAGGCCGGGGATTTATTTATTTAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA._T7_dsDNA_3	TAATACGAGTCACTATAGGTCCGCGGATTTATTTATTTAGTTTTAGAGCTAGAAATAGCA
fvvd_sgRNA._T7_dsDNA_4	TAATACGACTCACTATAGGTTCCGCGGATTTATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_5	TAATACGACTCACTATAGGATTCCGCGGATTTATTTATTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA._T7_dsDNA_6	TAATACGACTCACTATAGGTATTCCGCGGATTTATTTATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA._T7_dsDNA_7	TAATACGACTCACTATAGGTTATTCCGCGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA._T7_dsDNA_8	TAATACGACTCACTATAGGATTATTCCGCGGATTTATTTGTTTTAGAGCTAGAAATAGCA
fvvd_sgP.NA._T7_dsDNA_9	TAATACGACTCACTATAGGTATTATTGCGCGGATTTATTGTTTTAGAGCTAGAAATAGCA
1Svd_sgRNA._T7_dsDNA_10	TAATACGACTCACTATAGGATTATTATCCGCGGATTTATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_11	TAATACGACTCACTATAGGTATTATATTCCGCGGATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_12	TAATACGAGTCACTATAGGTTATTATATTCCGCGGATTTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_13	TAATACGACTCACTATAGGATTATTATATTCCGCGGATTGTTTTAGAGCTAGAAATAGCA
f-vd_sgRNA_77_dsDNA_14	FAATACGACTCACTATAGGTATTATTATATTCCGCGGATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_15	rAATACGACTCACTATAGGATTATTATTATTACCGCGGAGTTTTAGAGCTAGAAATAGCA
f..vci_sgRNA_77_dsDNA_18	FAATACGACTCACTATAGGATTATTATTATTATTACCGCGTTTTAGAGCTAGAAATAGCA
f¾d_sgRNA_77_dsDNA_noC	FAATACGACTCACTATAGGATATTAATTTATTTATTTAAGTTFTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_APOE4_C112R	FAATACGACTCACTATAGGGGAGGACGTGCGCGGCCGCCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_77_dsDNA_APOE4_C15SR	FAATACGACTCACTATAGGGAAGCGCCTGGCAGFGTACCGTTTTAGAGCTAGAAATAGCA
!wd_sgRNA_77_dsDNA_CTNNB1_T41A	FAATACGACTCACTATAGGCTGTGGCAGTGGCACCAGAAGTTTTAGAGCTAGAAATAGCA
f..vci_sgRNA_77_dsDNA_HRAS_Q61R	FAATACGACTCACTATAGGCCTCCCGGCCGGCGGTATCCGTTTTAGAGCTAGAAATAGCA
ftvd_sgRNA_T7_dsDNA_53_Y163C	FAATACGACTCACTATAGGGCTTGCAGATGGCCATGGCGGTTTTAGAGCTAGAAATAGCA
f-vd_sgRNA_T7_dsDNA_53_Y236C	FAATACGACTGACTATAGGACACATGCAGTTGTAGTGGAGTTTTAGAGCTAGAAATAGCA
f«¾_sciRNA_T7_dsDNA_53_N239D	FAATACGACTCACTATAGGTGTCACACATGTAGTTGTAGGTTTTAGAGCFAGAAATAGCA

[00441] Sequences of 80-nucleotide unlabeled strands and Cy3-labeled universal primer used in gel-based dsDNA deaminase assays. SEQ ID NOs: 224-248 appear from top to bottom below, respectively.

Cy3-primer	**Cy3-GTAGGTAGTTAGGATSAATGGAAGGTTGGTA**
dsDNA_2	GTCCATGGATCCAGAGGTCATCCATTAAATAAATAAATCCGCGGGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
<JsDNA_3	GTCCATGGATCCAGAGGTCATCCATAAATAAATAAATCCGCGGAAGCTATACCAACCTT<¾ATTCATCCTAACTACCTAC
dsDNA_4	GTCCATGGATCCAGAGGTCATCCATAATAAATAAATCCGCGGAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_5	GTCCATGGATCCAGAGGTCATCCAAATAAATAAATCGGCGGAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_6	GTCCATG^TCCAGAGGTCATCCMTAAATA^ CCGCGGAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_7	GTCCATGGATCCAGAGGTCATCCATAAATAAATCCGCGI¾WTAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_S	**GTCCATGGATCCAGAGGTCATO AAAATAAATCCGCGGAATAATGG^** TATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_9	GTCCATGGATCC^GAGGTCATCCAAATAAATCCGCGGAATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_10	**GTCCATGGATCCAGAGG7CA7CCAATAAA7CCGCGGATAATA&** TGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_11	GTCCATGGATCCAGAGGTCATCCATAAATCCGCGGAATATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
asDNA_12	**GTC<¾TGGATCCAGAGGT<IATCCAAAATCCGCGEAATATAATAAGGCTATACCAACCTRCCATTCATC^**TAACTACCTAC
SsDNA_13	GTCCATGGATCCAGAGGTCATCCAAATCCGCGGAATATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_14	GTCCATGGATCCAGAGGTCA TCCAATCCGCGGAATATAATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
asDNA_!5	GTCCATGGATCCAGAGGTCATCCATCCGCGGTAATAATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
asDNA_!S	GTCCATGGATCCAGAGGTCATCCAGCGGTAATAATAATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_noC	GTCCATGGATGCAGAGGTCATCCATTAAATAAATAAATTAATATTACTATACCAACCTTC€ATTCATCCTAACTACCTAC
dsDNA_8U	5CY3-G7AGGTAG7TASGATGAATGGAAGG77¾5TGTAGA7^TATCUGCGGATTTATTGGATGACCTCTGGATCCATGGACAT
dsDNA_APCE_C112R	**GCACCTCGCCGCGGTACTGCACC»GGCGGCCGCGCACGTCCTCCATG7CTACOWCCTTCCATTCATCCTAACTACCTAC**
asDNA_APOE_C15SR	CGGCGCCCTCGCGGGCCCCGGCCTGGTACACTGCCAGGCGCTTCTGCAGTACCAACCTTCCA7CATCCTAACTACCTAC
dsQNA_CTNN31_T41A	GTCTTACCTGGAGTCTGGAATCCATTCTGGTGC:CACTGCCACAGCTCCTTACCAACCTTCCATTCACCTAACTACCTAC
dsDNA_HRAS_Q61R	GGAGACGTGCCTGTTGGACATCCTGGATACCG<XGGCCGGGAGGAGTACTACCAACCTTCCATTCATCCTAACTACCTAC
asDNA_p53_Y1S3C	ACCCCCGCCCGGCACCCGCGTCCGCGCCATGGGCATCTGCAAGCAGTCATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_Y23SC	AGGTTGGCTCTGACTGTACCACCATCCACTACMCTGCATGTGTMCAGTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_N239D	TGGCTCTGACTGTA<XACCATCCACTACAACTACATGTGTGACAGTTCCTACCAACCTTCCATTCATCETAACTACCTAC

[00442] Primers used for generating PCR products to serve as substrates for T7 transcription of sgRNAs for high-throughput sequencing. rev_gRNA_T7 (above) was used in all cases. The pFYF1320 plasmid was used as template as noted in Materials and Methods section. SEQ ID NOs: 249-300 appear from top to bottom below, respectively.

fwd_sgRNA_T7_HTS_	base	TAA.TACGACTCACTATAGGTTATTTCGTGGATI ˙7AT7 TAG ˙TTT˙TAGAGC ˙AGAAATAGCA
fv,'d_sgRNA_T?_HTS_	1A	7AA7ACGACTCACTATAGGATATT ˙CGTGGA ˙TTATTTAG ˙˙˙˙TAGAGCTAGAAATAG CA
fvd_sgRNA_T7_**HTS**_	1C	TAATACGACTCACTATAGGCTATTTCGTGGAT1 ITATTTAG ˙TTTTAGAGC ˙AGAAATAGCA
fwd_sgRNA_T7_**HTS**_	1G	TAATACGACTCACTA7AGGGTATTTCG7GGAT ˙TATI TAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HT5_	2A	7AA TAGGAC7C AC7A7AGG7AAT77C GTGGA77TA77 T AG˙˙˙TTAGAG**CTAGAAATAG** CA
fwd_sgRNA_T?_**HTS**_	2C	7AA7 ACGAC7C AC7A7 AGG7C A777 CGT GGAT ˙TATTTAG ˙S˙TTAGAGCTAGAAATAGCA
f*vd_sg RNA_77_H7S_	2G	7AA7 ACGACTCACTA˙AGGTG ATTTC G7G GAT ITATTTAG˙˙˙TTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_	3T	7AATACGAC7CACTA7AGG77TT77CG7GGA77 TATT 7AG ˙TTTTAGAGCTAG AAA7.AGCA
fwd_sgRNA_T?_HTS_	3C	TAA7ACGAC7CAC7A7AGG77C777CG7GGA777A77 TAG ˙TT˙TAGAGCTAGAAATAGCA
fv,'d_3gRN**A_T?_HTS**_	3G	˙AA7 ACGAC7 CACTA 7AGG 7TGTTT CGT GGAT ˙TTA7 ˙TTAG˙1 ˙TTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_**HTS**_	4A	7AA7ACGAC7CAC7A7AGG77AA77CG7GGA71 I7A77TAG ˙TTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_**HTS**_	4C	TAATACGAC7CACTATAGGTTACTTCGTGGAT ˙TTATITAG ˙TTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T**?**_HTS_	4G	7AA7ACGAC7CAC7A7AGG77AG77CG7GGA7 ˙TTAT IT AGTTTTTAGAGCTAGAAA ˙TAGCA
fvvd_5gRNA_T7_HTS_	5A	7AA7ACGAC7CAC7A7AGG77A7A7CG7GGA71 **rTATTTAG** ˙S˙TTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_	**5C**	7AA7ACGAC7CAC7A7AGG77A7C7CG7GGAT I7A71TAG ˙TTTTAGAGC ˙AGAAATAGCA
fwd_sg**RNA_T7_HTS**_	5G	7AA7AC GAC7C AC7A7AG G77 A7G7C G7G GAT **rTATTTAGTTT** TAGAGCTAGAAATAGCA
f>vd_sgRMA_T7_HTS_	5A	7AATACGACTCAC7A7AGGTTATTACG7GGA7TTAT7 TAG ˙T/˙TTAGAGCTAGAAATAGCA
f*'d_sgRNA_T7 _**HTS**_	5C	˙AATACGAC7C ACTA7 AGG7TATTC CG7G GAT ITATTTAG ˙I˙(TTAGAGCTAGAAATAGCA
?vvd_sgRNA_T7_HTS_	**SG**	7AA7ACGAC7CAC7A7AGG77A77GCG7GGAT ˙TTAT1TAGTTTT AGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_	8A	7AATACGAC7C AC7A7A GG77A777 CA7G GA777 A7˙TAG ˙T˙/˙TAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_	8T	TAATAC GACTC ACTATAGG77ATTTCTT GGAT ˙TTA T7TAG ˙TTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7 _**HTS**_	8C	7**AATACG A**C7C AC7A7 AGGTTATTTG CTG GAT ITAT ˙TAGTT77AGAGC7AGAAA7AGCA
fwd_sgRN **A_T7_H** 7S_	9A	7AA7ACGAC7CAC7A7AGG77A777CGAGGAT TTAT˙T AGTTTTAGAGC ˙AGAAATAGCA
fwd_sgRNA_T7 _**HTS**_	9C	7AATACGAC7CAC7ATAGGTTAT7TCGCGGA7 7˙TA7T˙TAG7˙TTA GAGCTAG AAA 7AGCA
ftvd_sgRNA_T7_HTS_	SG	7**AA** TAG GAC7C ACTAT**AG** G77A 777CGGGGA77 TAT ˙TAGT˙˙˙TAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_	10A	7AA7ACGAC7CAC7A7AGG77A777CG7AGA777A7T T AGTTTTAGAGCT AGA**A**ATAGCA
fwci_sgRN**A**_T7_H7S_	10T	7AATAC GAC7C AC7A7AG GTTATTTC GT7GAT ITATTTAG ˙!˙TTT**AGAGCTAGAAATAGCA**
fwd_sgRNA_77_**HTS**_	10C	7AA7ACGAC7CAC7A7AGG77A777CG7CGA7 ˙I7AT ˙TAG ˙S˙TTAGAGCTAGAAATAGCA
f,v'd_sgRNA_T7_HTS_	11A	TAATACGACTCACTATAGGTTATTTCGTGAAT IT˙AT1TAG ˙T˙I˙TAGAGCTAGAAATAGCA
fwd_sgRNA_T7_**HTS**_	117	˙AATACGACTCAC7A7 AGG77ATTTCGTGTAT1˙7ATT˙AG777˙AGAGCTAG A**A**ATAGCA
fwd_sgRNA_T7_HTS_	11C	TAATAC GAC7C ACTA7AG GTTATTTC GTGC ATTTATT˙TAG˙T˙˙TTAGAGCTAGAAATAGCA
fwd_sgRNA_77_HTS_	12T	˙AATACG **A**CTCACTATAGGTTATTTCGTGG ˙T/˙s˙AT! TAG ˙T˙TTAGAGCTAGAAATAGCA
fwd_sgRNA_77 _**HTS**_	12C	TAATACGACTCACTATAGGTTATTTCGTGGC ˙T1˙AT ˙TA**G**T˙T˙TAGAGCTAGAAATAGCA

fwd_sgRNA_T7_HTS_ 12G	TAATACGACTCACTATAGGTTATTTCGTGGGTTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 13A	TAATACGACTCACTATAGGTTATTTCGTGGAATTA7TTAGTTTTAGAGCTAGAAATAGCA
fWd_sgRNA_77_H7S_ 13C	7AATACGACTCACTATAGGTTATTTCGTGG ACTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 13G	TAATACGACTCACTATAGGTTATTTCGTGGAGTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_77_HTS_ muitiC	TAATACGACTCACTATAGGTTCCCCCCCCGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fvd_sgRNA_T7_HTS_ TCGCACCC_odcS	TAATACGACTCACTATAGGCGCACCCGTGGATTTATTTAG7TTTAGAGCTAGAAATAGCA
fWd_sgRNA_77_HT3_ CCTCGCAC _o6d	TAATACGACTCACTATAGGCTCGCACGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fWd_sgRNA_77_HTS_ ACCCTCGC_odd	TAATACGACTCACTATAGGCCCTCGCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
ivd_sgRNA_T7_HTS_ GCACCCTC_o*1	TAATACGACTCACTATAGGCACCCTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ TCGCACCC_even	TAATACGACTCACTATAGGTCGCACCCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ CCTCGCAC _even	TAATACGACTCACTATAGGCCTCGCACGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS _ ACCCTCGC_sven	TAATACGACTCACTATAGGACCCTCGCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fWd_sgRNA_77_HT3. _ GCACCCTC_ever:	TAATACGACTCACTATAGGGCACCCTCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ EfMX1	TAATACGA CTCAC7ATAGG GAGTCC GAG CAGAAGAAGAAGTTTTAGAGCTAGA AATAGCA
fwd_sgRNA_77_HTS_ FANCF	TAATACGACTCACTATAGGGGAATCCCTTCTGCAGCACCGTTT7AGAGCTAGAAATAGCA
fwd_sgRNA_T7_H7S_ HEK293_site2	7AA7ACGACTCACTA7AGGGAACACAAAGCATAGACTGCGT7T7AGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ _ HEK2S3_site3	7AA7ACGACTCACTA7AGGGGCCCAGAC7GAGCACGTGAG7T7TAGAGCTAGAAA7AGCA
fwd_sgRNA_T7_HTS _ HEK293_site4	TAATACGACTCACTATAGGGGCACTGCGGCTGGAGGTGG3T7TTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS _ RNF2	TAATACGACTCAC7ATAGG GTCATC7T AG7C ATTACCTG GTTTTAGAGCTAGAAATAGCA

[00443] Sequences of *in vitro*-edited dsDNA for high-throughput sequencing (HTS). Shown are the sequences of edited strands. Reverse complements of all sequences shown were also obtained. dsDNA substrates were obtained by annealing complementary strands as described in Materials and Methods. Oligonucleotides representing the EMX1, FANCF, HEK293 site 2, HEK293 site 3, HEK293 site 4, and RNF2 loci were originally designed for use in the gel-based deaminase assay and therefore have the same 25-nt sequence on their 5'-ends (matching that of the Cy3-primer). SEQ ID NOs: 301-352 appear from top to bottom below, respectively.

Base sequence	ACGTAAACGGCCACAAGTTCTTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1A	ACGTAAACGGCCACAAGTTCATATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1C	ACGTAAACGGCCACAAGTTCCTATTTCGTGGATTTATTTATGGCATCTTCTTCA,AGGACG
1G	ACGTAAACGGCCACAAGTTCGTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
2A	ACGTAAACGGCCACAAGTTCTAATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
2C	ACGTAAACGGCCACAAGTTCTCATTTCGTGGATTTATTTATGGCATCTTCTTC AAGGACG
2G	ACGTAAACGGCCACAAGTTCTGATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3T	ACGTAAACGGCCACAAGTTCTTTTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3C	ACGTAAACGGCCACAAGTTCTTCTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3G	ACGTAAACGGCCACAAGTTCTTGTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4A	ACGTAAACGGCCACAAGTTCTTAATTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4C	ACGTAAACGGCCACAAGTTCTTACTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4G	ACGTAAACGGCCACAAGTTCTTAGTTCGTGGATTTATTTATGGCATCTTCTTCA¾GGACG
5A	ACGTAAACGGCCACAAGTTCTTATATCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5C	ACGTAAACGGCCACAAGTTCTTATCTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5G	ACGTAAACGGCCACAAGTTCTTATGTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
SA	ACGTAAACGGCCACAAGTTCTTATTACGTGGATTTATTTATGGCATCTTCTTCAAGGACG
SC	ACGTAAACGGCCACAAGTTCTTATTCCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
6G	ACGTAAACGGCCACAAGTTCTTATTGCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
8A	ACGTAAACGGCCACAAGTTCTTATTTCATGGATTTATTT ATGGCATCTTCTTCAAGGACG
8T	ACGTAAACGGCCACAAGTTCTTATTTCTTGGATTTATTTATGGCATCTTCTTCAAGGACG
8C	ACGTAAACGGCCACAAGTTCTTATTTCCTGGATTTATTTATGGCATCTTCTTCAAGGACG
9A	ACGTAAACGGCCACAAGTTCTTATTTCGAGGATTTATTTATGGCATCTTCTTCAAGGACG
SC	ACGTAAACGGCCACAAGTTCTTATTTCGCGGATTTATTTATGGCATCTTCTTCAAGGACG
3G	ACGTAAACGGCCACAAGTTCTTATTTCGGGGATTTATTTATGGCATCTTCTTC.AAGGACG
10A	ACGTAAAC GGCCACAAGTTC TTATTTCGTAGATTTATTTATGGCATCTTCTTC AAGGACG
10T	ACGTAAACGGCCACAAGTTCTTATTTCGTTGATTTATTTATGGCATCTTCTTCAAGGACG
10C	ACGTAAAC GGCCACAAGTTCTTATTTCGTC GATTTATTTATGGCATC TTCTTC AAGGACG
11A	ACGTAAACGGCCACAAGTTCTTATTTCGTGAATTTATTTATGGC ATCTTCTTCAAG GACG
11T	ACGTAAACGGCCACAAGTTCTTATTTCGTGTATTTATTTATGGCATCTTCTTCAAGGACG
11C	ACGT AAACGGCCACAAGTTCTTATTTCGTGCATTTATTT ATGGCATCTTCTTCAAGGACG

12T	ACGTAAACG GCCACAAGTTC TTATTTC GTG GTTTTATTT ATG GCATCTTCTTC AAGGACG
12C	ACGTAAACGGCCACAAGTTCTTATTTCGTGGCTTTATTTATGGCATCTTCTTCAAGGACG
12G	ACGTAAACGGCCACAAGTTCTTATTTCGTGGGT7ATTTATGGCATCTTCTTCAAGGACG
13A	ACGTAAACGGCCACAAGTTC TTATTTC GTG GAATTATTT ATG GCATC TTCTTC AAGGACG
13C	ACGTAAACGGCCACAAGTTCTTATTTCGTGGACTTATTTATGGCATCTTCTTC.AAGGACG
13G	ACGTAAACGGCCACAAGTTCTTATTTCGTGGAGTTATTTATGGCATCTTCTTCAAGGACG
muitiC	ACGTAAACGGCCACAAGTTCTTCCCCCCCCGATTTATTTATGGCATCTTCTTCAAGGACG
TCGCACCC _odd	ACGTAAACGGCCACAAGTTTCGCACCCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
CCTCGCAC_odd	ACGTAAACGGCCACAAGTTCCTCGCACGTGGATTTATTTATGGCATCTTCTTCAAGGACG
ACCCTCGC_odd	ACGTAA ACGGCCACAAGTT ACCC TCGCGTG GATTT ATTT ATGGCATCTT CTTC AAG GACG
GCACCCTCjJdd	ACGTAAACGGCCACAAGTTGCACCCTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
TCGCACCC _even	ACGTAAACGGCCACAAGTATTCGCACCCGTGGATTTATTATGGCATCTTCTTCAAGGACG
CCTCGCAC _even	ACGTAAACGGCCACAAGTATCCTCGCACGTGGATTTATTATGGCATCTTCTTCAAGGACG
ACCCTCGC_even	ACGT AAAC GGCCA CAAGTATACC CTC GCGTG GATTTATTATGGC ATCTT CTTC AAGGACG
GCACCCTC _8ven	ACGTAAACGGCCACAAGTATGCACCCTCGTGGATTTATTATGGCATCTTCTTCAAGGACG
EMX1 _invltro	GTAGGTAGTTAGGATGAATGGMGGTTGGTAGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTG
FANCFJ nvltro	**GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTCATGGAATCCTTCTGCAGCACCTGGATCGCTTTTCGAGCTTCTGG**
HEK293_s!te2._ invstro	**GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAACTGGAACACASAGCATAGACTGCGGGGCGGGCCAGCCTGAATAGCTG**
HEK2S3_S:te3._ inviiro	**GTAGGTAGTTAGGATGAATGGAAGGITGGTACTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAGCCCTGCT**
HEK293_s!te4_._ invitro	**GTAGGTAGTTAGGATGAATGGAAGGTTGGTACCGGTGGCACTGCGGOTGGAGGTGGGGGTTAAAGCGGAGACTCTGGTGC**
RNF2_inviiro	**GTAGGTAGTTAGGATGAATGGAAGGTTGGTATGGC AGTCATCTTAGTCATTACCTGAGGTGTTCGTT GTAACTCATATAA**

[00444] Primers for HTS of *in vitro* edited dsDNA. SEQ ID NOs: 353-361 appear from top to bottom below, respectively.

fwcU nVtro_HTS	ACACTC-TTTCC-CTACACGACGCTCTTCCGATCTNNNNACGTAAACGGCCACAA
rev_invitro_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGTCCTTGAAGAAGATGC
fvdj nvitfo_HEKJ afge!3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTAGGTAGTTAGGATGAATGGAA
rev_EMX1_!nviiro	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCGGTTGATGTGATGG
rev_FANCF_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGAAGCTCGGAAAAGC
rev_HEK2S3_3ite2_inviiro	TGGAGTTCAGACGTGTGCTCTTCCGATCTCAGCTATTCAGGCTGGC
r¾v_HEK2S3_Ste3J nvitro	TGGAGTTCAGA.CGTGTGCTCTTCC GATCTAG CAGGGCTTCCTTTC
rev_HEK293_Sfis4 _invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCACCAGAGTCTCCG
rev_RNF2_invitfo	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTATATGAGTTACAACGAACACC

[00445] Primers for HTS of on-target and off-target sites from all mammalian cell culture experiements. SEQ ID NOs: 362-469 appear from top to bottom below, respectively.

fwd_EMX1_HTS	**ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA**
rev_EMX1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
fwd_FANCF_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATTGCAGAGAGGCGTATCA
rev_FANCF_HT5	TGGAG7TCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGC7GAC
fwd_HEK233_site2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGCCCCATCTGTCAAACT
rev_HEK293_site2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAATGGATTCCTTGGAAACAATGA
fwd_HEK2S3_site3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAQAAAGG
rev_HEK2S3_site3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC
f,vd_HEK2S3_site4_HT3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAACCCAGGTAGCCAGAGAC
rev_HEK2S3_S:ts4_Hr3	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTTCAACCCGAACGGAG
f*vd_RNF2_HTS	**ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTCTTCTTTATTTCCAGCAATGT**
fsv_RNF2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTTTTCATGTTCTAAAAATGTATCCCA
fwd_p53_Y163C_HTS	**ACACTCTTTCCCTACACGACGCTCTTCCGATCTONNNTACAGTACTCCCCTGCCCTC**
rev_p53_Y163C_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATGTGCTGCTCACCATCGCTATCT
f*d_p53_N23SD_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTCATCTTGGGCCTGTGTT
rev_p53_N23SD_HT3	TGGAGTTCAGACGTGTGCTCTTCCGATCTAAATCGGTAA.GAGGTGGGCC
fwd_APOE4_C 15SR_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCGGACATGGAGGACGTG
rev_APOE4_C158R_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTTCCACCAGGGGCCC
fwd_EMX1_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGCCCAATCATTGATGCTTTT
rev_EMX1_off1_HTS	TGGAGTTCAG ACGTGTGCTCTTCC GATCTAGAAAC ATTTACCATAGACTATCAC CT
fwd_EMX1_off2_HTS	**ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGTAGCCTCTTTCTCAATGTGC**
rev_EMX1_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTCACAAGGATGCAGTCT
fwd_EMX1_off3_HTS	**ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTAGACTCCGAGGGGA**
fev_EMX1_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTCGTCCTGCTCTCACTT
fvvd_EMX1_off4_HTS	**ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGAGGCTGAAGAGGAAGACCA**
rev_EMX1_off4_HTS	TGGAGTT CAGACG TGTG CTC TTCC GATCTG GCCCAGCT GTGC A.7TCTAT

fwd_EMX1_off6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAAGAGGGCCAA.GTCCTG
rev_EMX1_oBS_HTS	TGGAGTTCAGACGTGTGCTCTTGCGATCTCAGCGAGGAGTGACAGCC
fwd_EMX1_off7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNHNCACTCCACCTGATCTCGGGG
fev_EMX1_off7_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGAG GAGGGAGGGAGCAG
fwd_EMX1_of?8_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCACAAATGCCCAAGAGAC
fev_EMX1_offS_HTS	TGGAGTTCAGACGTGTGCTCT7CCGATCTGACACAGTCAAGGGCCGG
fwd_EMX1_of?S_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCACCTTTGAGGAGGCAAA
rev_EMX1_offS_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCCATCTGAGAAGAGAGTGGT
fwd_EMX1_of?10_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCATACCTTGGCCCTTCCT
rev_EMX1_off10_HTS	TGGA.GTTCAGACGTGTGCTCTTCCGATCTTCCCTAGGCCCACACCAG
ftvd_FANCF_off1 _HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACCCACTGAAGAAGCAGGG
fev_FANCF_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCTTAATCCGGCTCCAT
fwd_FANCF_ofl2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCnviNNNTCCAGTGTTTCCATCCCGAA
rev_FANCF_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTGACCTCCACAACTCT
fwd_FANCF_cff3_HTS	**ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGGTACAGTTCTGCGTGT**
rev_FANCF_off3_HTS	TGGAGTTCAGACGTGTGCTCTTGCGATCTTCACTCTGAGCATCGCCAAG
fwd_FANCF_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNHNGGTTTAGAGCCAGTGAACTAGAG
rev_FANCF_o«4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAAGACAAAATCCTCTTTATACTTTG
fwd_FANCF_oS5_HT5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN^HNGGGAGGGGACGGCCTTAC
rsv_FANCF_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTCTGGCGAACATGGC
fwd_FANCF_off6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNMTCCTGGTTAAGAGCATGGGC
fev_FANCF_offS_HT3	TGGAGTTCAGACGTGTGCTCTTCCGATCTGATTGAGTCCCCACAGCACA
fwd_FANCF_off7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGTGTTTCCCATCCCCAA
rev_FANCF_off7_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGACCTCCACAACTGGAAAAT
fwd_FANCF_off8_HTS	ACACTC7TTCCCTACACGACGCTCTTCCGATCTNNNNGCTTCCAGACCCACCTGAAG
rsv_FANCF_offS_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCGAGGAAAATTGCTTGTCG
fwd_HEK2S3_site2_Off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGTGGAGAGTGAGTAAGCCA
rev_HEK233_site2_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACGGTAGGATGATTTCAGGCA
fwd_HEK2S3_site2_cff2_HTS	**ACACTCTTTCCCTACACGACGCTCTTCCGATCTNMNNCACAAAGCAGTGTAGCTCAGG**
fev_HEK2S3_site2_cfff2_H7S	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTTGGTACTCGAGTGTTATTCAG
fwd_HEK2S3_site3_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCCCTGTTGACCTGGAGAA
fev_HEK2S3_site3_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCCCTGACCA
fwd_HEK2S3_site3_cff2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNf-f!TGGTGTTGACAGGGAGCAA
rev_HEK2S3_site3_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGCAGAAGGG
twd_HEK233_site3_o?J3_HTS	ACAC7C7TTCCC7ACACGACGC7C77CCGA7C7NNNN7GAGAGGGAACAGAAGGGC7
fev_HEK2S3_site3_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCCAAGAACCT
fwd_HEK233_site3_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTAGCACTTTGGAAGGTCG
ev_HEK233_3ite3_off4_H7S	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCATCTTAATCTGCTCAGCC
fwd_HEK293_site3_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAAAGGAGCAGCTCTTCCTGG
fev_HEK2S3_sites_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGCACCATCTCCCACAA

fwd_HEK293_sits4_oS1_HTS	ACACTCTTTCCCTACACGACCeCTCTTCCGATCniN^GGCATSSCTTCTGAGACTCA
rev_HEK2S3_Site4_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTCCCTTGCACTCCCTGTCTTT
ftvd_HEK293_sits4_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCeATCTNNNNTTTGGCAATGGAGGCATTSG
rev_HEK2S3_Ste4_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAASAGGCTGCCCATGAGAG
fwd_HEK2S3_site4_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTCTGAGGCTCGAATCCTG
rev_HEK233_site4_sff3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTGGCCTCCATATCCCTG
fYf_HEK2S3_site4_o«4_HTS	ACACTCrrTTCCCTACACGACGCTCTTCCGATCTNN^-nTTCCACCAGAACTCAGCCC
r¾v_HEK2S3_Ste4_o«4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCGGTTCCTCCACAACAC
fwd_HEK2S3_site4_of5_HTS	AC^CTCTTTCCCTAaACGACGCTCTTCCGATCTNNNNCACGGGAAGGACAGGAGAAC
rev_HEK233_site4_cff5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGGGGAGGGATAAAGCAG
ftvd_HEK2S3_sit64_of»_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCT^NNNGCACGGGAGATGGCTTATGT
fsv_tt_EK2S3_3is4_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACA7C-CTCACTGTGCCACT
f*'d_HEK2S3_sits4_off7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCAGTCTCGGCCCCTCA
revJ HEK233_Site4_off7_H7S	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCACTGTAAAGCTCTTGGG
fwci_HEK233_sit64_off8_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTr^JNNAGGGTAGAGGGACAGAGCTG
rev_HEK293_site4_off8_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGACCCCACATAGTCAGTGC
fwd_HEK233_site4_off9_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNMNNGCTGTCAGCCCTATCTCCATC
rev_HEK2S3_Site4_off9_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGGGCAATTAGGACAGGGAC
fwd_HEK233_site4_off10_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCAGCGGAGGAGGTAGATTG
rev_HEK253_site4_off10_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCAGTACCTGGAGTCCCGA
fwd_HEK2_ChIP._off1._HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGACAGGCTCAGGAAAGCTGT
rev._HEK2_ChIP._off1._HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACACAAGCCTTTCTCCAGGG
fwd._HEK2_ChIP._off2._HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAATAGGGGGTGAGACTGGGG
rev._HEK2_ChIP_off2._HTS	TGGAGTTCAGACGT6TGCTCTTCCGATCTGCCTCAGACGAGACTTGAGG
fwd._HEK2__ChiP._off3._HTS	ACACTCTTTCCCTACACGACGGTCTTCCGATCTONNNGGCGAGCAGGAAAGGAATGT
rev HLK2 ChIP off3 HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGACTGCACCTGTAGCCATG
fwd._HEK2_chip_off4._HTS	ACACTCTTTCGCTACACGACGCTCTTCCGATCTNNNNTCAAGGAAATCACCCTGCCC
rev._HEK2_ChIP_off4._HTS	TGGAGTTCAGACGTGTGCTCTTGCGATCTAACTTCCTTGGTOTGCAGCT
fwd._HEK2._ChIP__off5__HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTMNNNATGGGCTCAGCTACGTCATG
rev HEK2 ChIP of 5 HTS	TGGAGTTCAGACGTGTGGTCTTCCGATCTAATAGCAGTGTGGTGGGCAA
fwd._HEK3._ChIP._off1._HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGCACATCCCTTGTCTCTCT
rev._HEK3__ChIP._off1._HTS	TGGAGTTC AGAC GTGTG CTGTTC CGATCTCTACTGG AGCAC ACC CCAAG
fwd._HEK3._ChiP._off2._HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGGGTCACGTAGCTTTGGTC
rev._HEK3_Chip._off2._HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGGTGGCCATGTGCAACTAA
fwd._HEK3._ChiP._off3._HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTACTACGTGCCAGCTCAGG
rev__HEK3__ChIP._off3._HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCTCCCCTCCTCACTAACC
fwd._HEK3._ChiP._off4._HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCTCAGCTCCATTTCCTGT
rev._HEK3_.ChiP._cff4_.HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAAGCTTTATGGCACCAGGGG
fwd._HEK3._ChIP._Qff6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTCAGCATTAGCAGGCT
rev._HEK3._Chip._off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCCTGGCTTTCCGATTCCC

fwd_HEK4_ChiP_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGCAATTGGAGGAGGAGCT
rev_HEK4_ChIP_offi_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCAGCTACAGGCAGAACA
fwd_HEK4_Ch!P_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTACCCCCAACACAGATGG
rev_HEK4_Chip_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCACACAACTCAGGTCCTCC

[00446] Sequences of single-stranded oligonucleotide donor templates (ssODNs) used in HDR studies.

EMX1 sense (SEQ ID NO: 470)

TCATCTGTGCCCCTCCCTCCCTGGCCCAGGTGAAGGTGTGGTTCCAGAACCGGAGGACAAAGTACA
AACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTTTGAGCAGAAGAAGAAGGGCTCCCATCACATC
AACCGGTGGGGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACCTCCAATGACTAG
GGT

EMX1 antisense (SEQ ID NO: 471)

ACCCTAGTCATTGGAGGTGACATCGATGTCCTCCCCATTGGCCTGCTTCGTGGCAATGCGCCACCG
GTTGATGTGATGGGAGCCCTTCTTCTTCTGCTCAAACTCAGGCCCTTCCTCCTCCAGCTTCTGCCGT
TTGTACTTTGTCCTCCGGTTCTGGAACCACACCTTCACCTGGGCCAGGGAGGGAGGGGCACAGATG
A

HEK293 site 3 sense (SEQ ID NO: 472)

CATGCAATTAGTCTATTTCTGCTGCAAGTAAGCATGCATTTGTAGGCTTGATGCTTTTTTTCTGCTTCT
CCAGCCCTGGCCTGGGTCAATCCTTGGGGCTTAGACTGAGCACGTGATGGCAGAGGAAAGGAAGC
CCTGCTTCCTCCAGAGGGCGTCGCAGGACAGCTTTTCCTAGACAGGGGCTAGTATGTGCAGCTCCT

HEK293 site 3 antisense (SEQ ID NO: 473)

AGGAGCTGCACATACTAGCCCCTGTCTAGGAAAAGCTGTCCTGCGACGCCCTCTGGAGGAAGCAGG
GCTTCCTTTCCTGTGCCATCACGTGCTCAGTCTAAGCCCCAAGGATTGACCCAGGCCAGGGCTGGA
GMGCAGAAAAAAAGCATCAAGCCTACAAATGCATGCTTACTTGCAGCAGAAATAGACTAATTGCATG

HEK site 4 sense (SEQ ID NO: 474)

GGCTGACAAAGGCCGGGCTGGGTGGAAGGAAGGGAGGAAGGGCGAGGCAGAGGGTCCAAAGCAG
GATGACAGGCAGGGGCACCGCGGCGCCCCGGTGGCATTGCGGCTGGAGGTGGGGGTTAAAGCGG
AGACTCTGGTGCTGTGTGACTACAGTGGGGGCCCTGCCCTCTCTGAGCCCCCGCCTCCAGGCCTGT
GTGTGT

HEK site 4 antisense (SEQ ID NO: 475)

ACACACACAGGCCTGGAGGCGGGGGCTCAGAGAGGGCAGGGCCCCCACTGTAGTCACACAGCACC
AGAGTCTCCGCTTTAACCCCCACCTCCAGCCGCAATGCCACCGGGGCGCCGCGGTGCCCCTGCCT
GTCATCCTGCTTTGGACCCTCTGCCTCGCCCTTCCTCCCTTCCTTCCACCCAGCCCGGCCTTTGTCA
GCC

APOE4 sense (SEQ ID NO: 476)

AGCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCG
CGATGCCGATGACCTGCAGAAGTGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAG
CGCGGCCTCAGCGCCATCCGCGAGCGCCTGGGGCCCCTGGTGGAACAGGGCCGCGTGCGGGCCG
CCACTGT

APOE4 antisense (SEQ ID NO: 477)

ACAGTGGCGGCCCGCACGCGGCCCTGTTCCACCAGGGGCCCCAGGCGCTCGCGGATGGCGCTGA
GGCCGCGCTCGGCGCCCTCGCGGGCCCCGGCCTGGTACACTGCCAGGCACTTCTGCAGGTCATCG
GCATCGCGGAGGAGCCGCTTACGCAGCTTGCGCAGGTGGGAGGCGAGGCGCACCCGCAGCTCCT
CGGTGCT

p53 Y163C sense (SEQ ID NO: 478)

ACTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTGA
TTCCACACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCACAGCACATGAC
GGAGGTTGTGAGGCGCTGCCCCCACCATGAGCGCTGCTCAGATAGCGATGGTGAGCAGCTGGGGC
TG

p53 Y163C antisense (SEQ ID NO: 479)

CAGCCCCAGCTGCTCACCATCGCTATCTGAGCAGCGCTCATGGTGGGGGCAGCGCCTCACAACCTC
CGTCATGTGCTGTGACTGCTTGTAGATGGCCATGGCGCGGACGCGGGTGCCGGGCGGGGGTGTGG
AATCAACCCACAGCTGCACAGGGCAGGTCTTGGCCAGTTGGCAAAACATCTTGTTGAGGGCAGGGG
AGT

[00447] Deaminase gene gBlocks Gene Fragments

hAlD (SEQ ID NO: 169)

rAPOBECl (mammalian)(SEQ ID NO: 170)

CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGAGCTCAGAGACTGGCCCAGTGGCTGTGGA .ee
CCACATTGAGACGGCGGATCGAGCCCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCA
AGGAGACCTGCCTGCTTTACGAAATTAATTGGGGGGGCCGGCACTCCATTTGGCGACATACATCACA
GAACACTAACAAGCACGTCGAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTCCG
AACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAGTAGGGCCATC
ACTGAATTCCTGTC.AAGGTATCCCCACGTCACTCTGTTTATTTACATCGCAAGGCTGTACCACCACGC
TGACCCCCGCAATCGACAAGGCCTGCGGGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACT
GAGCAGGAGTCAGGATACTGCTGGAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGG
CCTAGGTATCCCCATCTGTGGGTACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGC
CTCCTTGTCTCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCT
TGTCATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAATGAGCGGCCGCTCGA
TTGGTTTGGTGTGGCTCTAA

pmCDAl (SEQ ID NO: 171)

CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGACAGACGCTGAATATGTTAGGATCCATGAAAA
ACTGGATATCTATACATTTAAGAAGCAGTTCTTCAATAACAAAAAGTCAGTATCTCACAGATGCTATGT
CCTGTTCGAACTCAAGAGAAGAGGAGAAAGGCGGGCCTGTTTCTGGGGGTACGCGGTTAATA.AACC
CCAGTCCGGGACCGAGAGGGGGATTCACGCCGAGATCTTTTCAATTAGGAAGGTTGAAGAGTATCT
TCGCGACAATCCCGGTCAGTTCACAATTAACTGGTACAGCTCCTGGAGCCCTTGCGCTGATTGCGCC
GAGAAAATACTCGAATGGTACAACCAGGAGTTGAGAGGCAATGGCCACACTCTCAAGATTTGGGCTT
GCMGCTTTACTACGAGAAGAACGCGAGAAATCAGATTGGCTTGTGGAACCTCAGGGACAACGGGG
TCGGGTTGAATGTTATGGTGTCCGAACATTACCAGTGCTGTAGAAAGATCTTCATTCAGTCCAGTCAC
AATCAGCTGAACGAGAACAGATGGCTGGAGAAAACACTGAAACGGGCAGAGAAAAGGCGCTCAGAG
CTGAGTATCATGATCCAGGTCAAAATCCTGCATACAACCAAAAGCCCGGCTGTATAAGCGGCCGCTC
GATTGGTTTGGTGTGGCTCTAA

haPOBEC3G (SEQ ID NO: 172)

GATCCTTGGTACCGAGCTCGGATCCAGGCACCATGGAGCTGAAGTATCACCCTGAGATGGGGTTTTT
CCACTGGTTTAGTAAGTGGCGCAAACTTCATCGGGATCAGGAGTATGAAGTGACCTGGTATATCTCT
TGGTCTCCCTGCACAAMTGTACACGCGACATGGGCACATTTCTGGCCGAGGATCCAAAGGTGACG
CTCACAATCTTTGTGGCCCGCCTGTATTATTTCTGGGACCCGGATTATCAGGAGGCACTTAGGTCAT
TGTGCCA.AAAGCGCGACGGACCACGGGCGACTATGAAAATCATGAATTATGACGAATTCCAGCATTG
CTGGAGTAAGTTTGTGTACAGCCAGCGGGAGCTGTTCGAGCCCTGGAACAATCTTCCCAAGTACTAC
ATACTGCTTCACATTATGTTGGGGGAGATCCTTCGGCACTCTATGGATCCTCCTACCTTTACGTTTAA
CTTTAATAATGAGCCTTGGGTTCGCGGGCGCCATGAAACCTATTTGTGCTACGAGGTCGAGCGGATG
CATAATGATACGTGGGTCCTGCTGAATCAGAGGAGGGGGTTTCTGTGTAACCAGGCTCCACATAAAC
ATGGATTTCTCGAGGGGCGGCACGCCGMCTGTGTTTCCTTGATGTGATACCTTTCTGGAAGCTCGA
CCTTGATCAAGATTACAGGGTGACGTGTTTCACCTCCTGGTCACCCTGCTTCAGTTGCGCCCAAGAG
ATGGCTAAATTTATCAGTAAGAACAAGCATGTGTCCCTCTGTATTTTTACAGCCAGAATTTATGATGAC
CAGGGCCGGTGCCAGGAGGGGCTGCGGACACTCGCTGAGGCGGGCGCGAAGATCAGCATAATGA
CATACTCCGAATTCAAACACTGTTGGGACACTTTTGTGGACCACCAGGGCTGCCCATTTCAGCCGTG
GGATGGGCTCGACGAACATAGTCAGGATCTCTCAGGCCGGCTGCGAGCCATATTGCAGAACCAGGA
GAATTAGGCGGCCGCTCGATTGGTTTGGTGTGGCTCTAA

rAPOBECl(E. Coli) (SEQ ID NO: 173)

GGCCGGGGATTCTAGAMTAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGATGTCTTCTGAAA
CCGGTCCGGTTGCGGTTGACCCGACCCTGCGTCGTCGTATCGAACCGCACGAATTCGAAGTTTTCT
TCGACCCGCGTGAACfTGCGTAAAGAAACCTGCCTGCTGTACGAAATCAACTGGGGTGGTCGTCACT
CTATCTGGCGTCACACCTCTCAGAACAGCAACAAACACGTTGAAGTTAACTTCATCGAAAAATTCACC
ACCGAACGTTACTTCTGCCCGAACACCCGTTGCTCTATCACCTGGTTCCTGTCTTGGTCTCCGTGCG
GTGAATGCTCTCGTGCGATCACCGAATTCCTGTCTCGTTACCCGCACGTTACCGTGTTCATCTACATC
GC GCGTCTGT ACC ACC ACG CGG AC CCG CGTAACC GTC AGGGTCTGCGTG ACCTG ATCTCTTCTGGT
GTTACCATCCAGATCATGACCGAACAGGAATCTGGTTACTGCTGGCGTAACTTCGTTAACTACTCTCC
GTCTAACGAAGCGCACTGGCCGCGTTACCCGCACCTGTGGGTTCGTCTGTACGTTCTGGAACTGTA
CTGCATCATCCTGGGTCTGCCGCCGTGCCTGAACATCCTGCGTCGTAAACAGCCGCAGCTGACCTT

CTTCACCATCGCGCTGCAGTCTTGCCACTACCAGCGTCTGCCGCCGCACATCCTGTGGGCGACCGG
TCTGAAAGGTGGTAGTGGAGGGAGCGGCGGTTCAATGGATAAGAAATAC

[00448] Amino Acid Sequences of NBE1, NBE2, and NBE3.

NBE1 for *E. Coli* expression (His_6-rAPOBECl-XTEN-dCas9) (SEQ ID NO: 154)

MGSSHHHHHMSSETGPVAVDPTLRRRiEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNF!EKFTTERYFCPNTRCSfTWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGL
RDL!SSGVTiQ!MTEQESGYCWRNFVNYSPSNEAHWPRYPHLVWRLYVLELYCfiLGLPPCLN!LRRKQPQ
LTFFTiALQSCHYQRLPPHiLWATGLKSGSETPGTSESATPESDKKYSiGLAiGTNSVGVVAVlTDEYKVPSK
KFKVLGNTDRHSiKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRiCYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDK|KHERHPiFGNiVDEVAYHEKYPTiYHLRKKLVDSTDKADLRLIYLALAHy!KFRGHFL!EGDL
NPDNSDVDKLF!!QLVQTYNQLFEENPiNASGVDAKA!LSARLSKSRRLENUAQLPGEKKNGLFGNLlALSL
GLTPNFKSNFDLAEOAKLQLSKDTYDDDLDNLLAQ!GDQYADLFLAAKNLSDA!LLSDILRVNTEITKAPLSA
SMlKRYDEHHQDLTliKALVRQQLPEKYKE!FFDQSKNGYAGYIDGGASQEEFYKFIKP!LEKMDGTEELL
VKLNREDLLRKQRTFDNGS!PHQiHLGELHAILRRQEDFYPFLKDNREKIEK!LTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEWDKGASAQSFlERMTNFDiWLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKK!ECFDSVEiSGVEDRFNASLGTYHDLLK!tKDK
DFLDNEENEDILEDIVLTLTLFEDREMiEERLKTY'AHLFDDKVMKQLKRRRYTGWGRLSRKLSNGIRDKQS
GKTILDFLKSDGFANRNFMGLIHDDSLTFKEDlQKAQVSGQGDSLHEHiANLAGSPAiKKGILQTVKVVDEL
VKVMGRHKPENJVIEMARENQTTQKGQKNSRERMKRIEEGiKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDS1DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFiKRQLVETRQITKHVAQILDSRlVlNTKYDENDKLIREVKViTLK
SKLVSDFRKDFQR'KVREiNNYHHAHDAYLNAVVGTALiKK>'PKLESEFVYGDYK\A'DVRKMIAKSEQEiG
KATAKYFFYSNIMNFFKTEiTLANGEiRKRPLiETNGETGEIV\ADKGRDFATVRKVl-SMPQW iVKKTEVQT
GGFSKESILPKRNSDKLlARKKDWDPKKYGGFDSPTVAYSVLWAKVEKGKSKKLKSVKELLGITiMERSS
FEKNPiDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEHEQ!SEFSKRViLADANLDKVLSAYNKHRDKP!REQAEN!IHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLiHQS!TGLYETRIDLSQLGGDSGGSPKKKRKV

NBE1 for Mammalian expression (rAPOBECl-XTEN-dCas9-NLS) (SEQ ID NO: 155)

MSSETGPVAVDPTLRRRiEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRA!TEFLSRYPHVTLFIYiARLYHHADPRNRQGLRDLISSGVT!Q
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYC!ILGLPPCLNiLRRKQPQLTFFTIALQSC
HYQRLPPHiLWATGLKSGSETPGTSESATPESDKKYSIGLAtGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HStKKNL!GALLFDSGETAEAl†¾LKRTARRRYTRRKNR!CYLQEiFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHP!FGN!VDEVAYHEKYPT!YHLRKKLVDSTDKADLRLiYLALAHM!KFRGHFLIEGDLNPDN!SDVDKLF
IQLVQTYNQLFEENPiNASGVDAKA!LSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSHFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDA!LLSD!LRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEiFFDQSKNGYAGYiDGGASQEEFYKFSKPiLEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQ!HLGELHA!LRRQEDFYPFLKDNREK!EKILTFRIPYYVGPLARGNSRFAWfv!TRKSEETITP
WNFEEVVDKGASAQSFSERMTNFDKNLPNEKVLPKHSLLYEYFTV^'NELTKVKYVTEGMRKPAFLSGEQ
KKAiVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKiiKDKDFLDNEENEDILE
D!VLTLTLFEDREMiEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING!RDKQSGKT!LDFLKSDGFA
NRNFMQLIHDDSLTFKEDiQKAQVSGQGDSLHEHiANLAGSPAfKKGILQTVKVVDELVKVMGRHKPENiVI
EMARENQTTQKGQKNSRERMKRiEEG!KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAiVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEWKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGF!KRQLVETRQiTKHVAQiLDSRMNTKYDENDKLIREVKV!TLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNiM
NFFKTE!TLANGEiRKRPLiETNGETGEiVWDKGRDFAWRKVLSMPQVNIVKKTEVQTGGFSKES!LPKRN
SDKL!ARKKDWDPKKYGGFDSPTVAYSVLWAKVEKGKSKKLKSVKELLGiTiMERSSFEKNPIDFLEAKG
YKEVKKDLilKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDESIEQiSEFSKRVILADANLDKVLSAYHKHRDKPiREQAEN!IHLFTLTiNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLSHQS!TGLYETRIQLSQLGGDSGGSPKKKRKV

Alternative NBE1 for Mammalian expression with human APOBEC1 (hAPOBEC 1-XTEN-dCas9-NLS) (SEQ ID NO: 158)

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNT
TNHVEVNFIKKFTS ERDFHPS MSCSITWFLS WSPCWECS QAIREFLS RHPG VTLVIY VA
RLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPP
LWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPS
VAWRGS ETPGTS E SATPES DKKYS IGLAIGTNS VGW A VITDE YKVPS KKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH
RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL
AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIALS LGLTPNFKS NFDLAED AKLQLS KDTYDDDL
DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL
LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK
LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLP
KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL
KSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV

KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKE
HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK
VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKD
FQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSrrGLYETRIDLSQLGGDSGGSPKKK
RKV

NBE2 (rAPOBECl-XTEN-dCas9-UGI-NLS) (SEQ ID NO: 156)

MSSETGPVAVDPTLRRREPHEFEVFFDPRELRKETCLLYESNWGGRHSIWRHTSQNTNKHVEVNFiEKF
TTERYFCPNTRCSSTWFLSWSPCGECSRA!TEFLSRYPHVTLF!YiARLYHHADPRNRQGLRDLISSGVT!Q
jMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSiGLA!GTNSVGWAVITOEYKVPSKKFKVLGNTDR
HSIKKNL!GALLFDSGETAEATRLKRTARRRYTHRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNiVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL!YLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
iQLVQTYNQLFEENPiNASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQiGDQYADLFLAAKNLSDAILLSD!LRWlTEJTKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEi-LVKLNREDLLRKQ
RTFDNGSfPHQIHLGELHAiLRRQEDFYPFLKDNREK!EKlLTFRIPYYVGPLARGNSRFAWMTRKSEET!TP
WNFEEWDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEfSGVEDRFNASLGTYHDLLKi!KDKDFLDNEENEDiLE
DIVLTLTLFEDREMiEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTiLDFLKSDGFA
NRNFMQLiHDDSLTFKEDiQKAQVSGQGDSLHEHIANLAGSPAlKKGILQTVKWDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGiKELGSQtLXEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAlVPQSFLKDDSiDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKL!TQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQ!TKHVAQiLDSRMNTKYDENDKLlREVKViTLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAWGTALlKKYPKLESEFWGDYKVYDVRKMIAKSEQE!GKATAKYFFYSNfM
NFFKTEITLANGEIRKRPUETNGETGEiVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKL!ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG!TiMERSSFEKNPlDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEHEQiSEFSKRViLADAN!LDKVLSAYNKHRDKPiREQAENIiHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLiHQSiTGLYETRiDLSQLGGDSGGSTNLSDiiEKETGKQLViQESiLMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPVVALV!QDSNGENKIKMLSGGSPKKKRKV

NBE3 (rAPOBECl-XTEN-Cas9n-UGI-NLS) (SEQ ID NO: 157)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHS!VVRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSVVSPCGECSRA!TEFLSRYPHVTLF!YIARLYHHADPRNRQGLRDLISSGVT!Q
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYC!!LGLPPCLN!LRRKQPQLTFFTIALQSC
IHYQRLPPHFLWATGLKSGSETPGTSESATPESDKKYS!GLAIGTNSVGWAVITDE\'KVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRSCYLQEIFSNEfvIAKVDDSFFHRLEESFLVEEDKK
HERHP!FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLTYLALAHM!KFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVT>AKAILSARLSKSRRL£NUAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKMLSDAILLSD!LRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKE!FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGS!PHQ!HLGELHAILRRQEDFYPFLKDNREKIEKILTFR!PYYVGPLARGNSRFAWMTRKSEET!TP
WNFEEWDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKK!ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
D!VLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING!RDKQSGKTILDFLKSDGFA
NRNFMQLFHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGSLQWKWDELVKVMGRHKPENIVI
EMARENQTTQKGQ!WSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLS DYDVDHIVPQS FLKDDS IDNKVLTRS DKNRGKS DNVPS EEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREtNNYHHAHDAYLNAWGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQE!GKATAKYFFYSNIM
NFFKTE!TLANGEIRKRPLIETNGETGEIW/DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLWAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLHKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLFHQS!TGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

pmCDAl-XTEN-dCas9-UGI (bacteria) (SEQ ID NO: 159)
MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKP
QSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGN
GHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQL
NENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPESDKKYSI
GLAIGTNS VGW A VITDE YKVPS KKFKVLGNTDRHS IKKNLIG ALLFDS GETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQT YNQLFEENPIN A SGVDAKAILS ARLS KSRRLENLIAQLPGEKKNGLFGNL
IALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLS GEQKKAIVDLLFKTNRKVT VKQLKED YFKKIECFDS VEIS GVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED
IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR
ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES
ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQS ITGLYETRIDLS QLGGDS GGSMTNLSDIIEKETGKQLVIQES ILMLPEE VEE
VIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML

pmCDAl-XTEN-nCas9-UGI-NLS (mammalian construct) (SEQ ID NO: 160)

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKP
QSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGN
GHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQL
NENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPESDKKYSI
GLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNL
IALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED
IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR
ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES
ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI
GNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPK
KKRKV

huAPOBEC3G-XTEN-dCas9-UGI (bacteria) (SEQ ID NO: 161)
MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGF
LEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIF
TARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEH
SQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM
AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGAS QEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNS RFAWMTRKS EETITPWNFEE VVDKGASAQSFIERMTNFDK
NLPNEKVLPKHS LLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYS LFELENGRKRMLAS AGELQKGNELALPS KYVNFLYLAS HYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGDSGGSMTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDE
NVMLLTSDAPEYKPWALVIQDSNGENKIKML

huAPOBEC3G-XTEN-nCas9-UGI-NLS (mammalian construct) (SEQ ID NO: 162)

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGF
LEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIF
TARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEH
SQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM
AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGAS QEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIP YYVGPLARGNS RFAWMTRKS EETITPWNFEE VVDKGASAQSFIERMTNFDK
NLPNEKVLPKHS LLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYS LFELENGRKRMLAS AGELQKGNELALPS KYVNFLYLAS HYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDEN
VMLLTSDAPEYKPWALVIQDSNGENKIKMLS GGSPKKKRKV

huAPOBEC3G (D316R_D317R)-XTEN-nCas9-UGI-NLS (mammalian construct) (SEQ ID NO: 163)

**MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGF
LEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIF
TARIYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEH
SQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM
AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS QEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGLRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLLREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGELRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPLOFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS HYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDEN
VMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV**

[00449] **Base Calling Matlab Script**

WTnuc='GCGGACATGGAGGACGTGCGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCG
GCCAGA
GCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGC
CGATGAC
CTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAGCGCGGCCTCAGCGCCATC
CGCGAGCG CCTGGGGCCCCTGGTGGAACAG'(SEQ ID NO: 164);

```
%cycle through fastq files for different samples files=dir('*.fastq');
for d= 1:20
filename=files(d).name;
%read fastq file
[header,seqs,qscore]=fastqread(filename);
seqsLength = length(seqs);                                  % number of sequences seqsFile=
strrep(filename,'. fastq','');                              % trims off .fastq
%create a directory with the same name as fastq file ifexist(seqsFile,'dir');
error('Directory already exists. Please rename or move it before moving on.');
end
mkdir(seqsFile);                                            % make directory
wtLength = length(WTnuc);                                   % length of wildtype sequence
%% aligning back to the wildtype nucleotide sequence
%
```

```
% AIN is a matrix of the nucleotide alignment  window=l :wtLength;
sBLength = length(seqs);                                  % number of sequences
% counts number of skips  nSkips = 0;
ALN=repmat(' ',[sB Length wtLength]);
% iterate through each sequencing read  for i = 1:sBLength
%If you only have forward read fastq files leave as is
%If you have R l foward and R2 is reverse fastq files uncomment the
%next four lines of code and the subsequent end statement
%                 ifmod(d,2)==0;
%                         reverse = seqrcomplement(seqs  {i});
%                         [score, alignment, start] =
swalign(reverse,WTnuc,Alphabet','NT');
%                 else

[score,alignment,start]=swalign(seqs{i},WTnuc,'  Alphabet', 'NT');
%                 end

% length of the sequencing read  len=
length(alignment(3 ,:));
% if there is a gap in the alignment , skip = 1 and we will
% throw away the entire read  skip = 0;
for j = l :len
if (alignment(3,j) == '-' ,, alignment(l,j)  == '-')  skip = 1;
                          break;
end
%in addition if the qscore for any given base in the read is
%below 31 the nucleotide is turned into an N (fastq qscores that  are not letters)
ifisletter(qscore  {i}(start( 1)+j - 1)) else
alignment(l,j)  = 'N;
end

end
if skip == 0 && len>10
ALN(i,    start(2):(start(2)+length(alignment)-l))=alignment(l,:);
```

```
        end
end
```

```
% with the alignment matrices we can simply tally up the occurrences of
% each nucleotide at each column in the alignment these
% tallies ignore bases annotated as N
% due to low qscores
TallyNTD=zeros(5,wtLength); fori= 1:wtLength

TallyNTD(:,i)=[sum(ALN(:,i)=='A'),sum(ALN(:,i)=='C'),sum(ALN(:,i)=='G'),sum(A
LN(:,i)=='T'),sum(ALN(:,i)=='N')];
end
% we then save these tally matrices in the respective folder for
% further processing

save(strcat(seqsFile,7TallyNTD'), 'TallyNTD'); dlmwrite(strcat(seqsFile, '/TallyNTD.txt'), TallyNTD, 'precision',
'%.3f', 'newline', 'pc'); end
```

[00450] INDEL Detection Matlab Script

```
WTnuc='GCGGACATGGAGGACGTGCGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCGGC
CAGA
GCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGCCG
ATGAC
CTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAGCGCGGCCTCAGCGCCATCCGC
GAGCG CCTGGGGCCCCTGGTGG AACAG'(SEQ ID NO: 164);
%cycle through fastq files for different samples files=dir('*.fastq');
%specify start and width of indel window as well as length of each flank indelstart=154;
width=30; flank=10;

for d=l:3
filename=files(d).name;
%read fastq file
[header,seqs,qscore]=fastqread(filename);
seqsLength = length(seqs); % number of sequences seqsFile
=strcat(strrep(filename,'. fastq', ''),'_INDELS');
%create a directory with the same name as fastq file+_INDELS ifexist(seqsFile,'dir');
error('Directory already exists. Please rename or move it before moving on.');
end
mkdir(seqsFile); % make directory
wtLength = length(WTnuc); % length of wildtype sequence sBLength =
length(seqs); % number of sequences

% initialize counters and cell arrays
nSkips = 0; notINDEL=0;
ins={ };
dels={ }; Numlns=0;
NumDels=0;
% iterate through each sequencing read for i = 1:sBLength
%search for 10BP sequences that should flank both sides of the "INDEL WINDOW"
windowstart=strfind(seqs {i },WTnuc(indelstart-flank:indelstart));

windowend=strfind(seqs {i },WTnuc(indelstart+width:indelstart+width+flank
));
%if the flanks are found proceed
```

```
if length(windowstart)== 1&&length(windowend)== 1
%if the sequence length matches the INDEL window length save as
%not INDEL
if   windowend-windowstart==width+flank notINDEL=notINDEL+l  ;
%if the sequence is two or more bases longer than the INDEL
%window length save as an Insertion
elseif   windowend-windowstart>=width+flank+2 NumIns=NumIns+ 1;
ins {Numlns }=seqs {i };
%if the sequence is two or more bases shorter than the INDEL
%window length save as a Deletion
elseif  windowend-windowstart<=width+flank-2  NumDels=NumDels+l  ;
dels {NumDels }=seqs {i };
%keep track of skipped sequences that are either one base
%shorter or longer than the INDEL window width  else
nSkips=nSkips+ 1;
end
%keep track of skipped sequences that do not possess matching flank
%sequences else
nSkips=nSkips+ 1;

            end
end
```

```
fid=fopen(strcat(seqsFile, 7summary.txt'),'wt');
fprintf(fid, 'Skipped reads %i\n not INDEL %i\n Insertions %i\n Deletions
%i\n', [nSkips, notlNDEL, Numlns, NumDels]);   fclose(fid);
save(strcat(seqsFile,'/nSkips'),'nSkips');    save(strcat(seqsFile,7notINDEL'),'notINDEL');
save(strcat(seqsFile,7NumIns'), 'Numlns');   save(strcat(seqsFile,'/NumDels'),'NumDels');
save(strcat(seqsFile, '/dels'),'dels');
C = dels;
fid = fopen(strcat(seqsFile, '/dels.txt'), 'wt');   fprintf(fid, '"%s"\n', C{:});
fclose(fid);
save(strcat(seqsFile,'/ins'),'ins');    C = ins;
fid = fopen(strcat(seqsFile, '/ins.txt'), 'wt');   fprintf(fid, '"%s"\n', C{:});
fclose(fid);

end
```

EXAMPLE 5: Cas9 variant sequences

[00451] The disclosure provides Cas9 variants, for example Cas9 proteins from one or more organisms, which may comprise one or more mutations *{e.g.,* to generate dCas9 or Cas9 nickase). In some embodiments, one or more of the amino acid residues, identified below by an asterek, of a Cas9 protein may be mutated. In some embodiments, the D10 and/or H840 residues of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, are mutated. In some embodiments, the D10 residue of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, is mutated to any amino acid residue, except for D. In some embodiments, the D10 residue of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, is mutated to an A. In some embodiments, the H840 residue of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding residue in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, is an H. In some embodiments, the H840 residue of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, is mutated to any amino acid residue, except for H. In some embodiments, the H840 residue of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding mutation in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, is mutated to an

A. In some embodiments, the D10 residue of the amino acid sequence provided in SEQ ID NO: 6, or a corresponding residue in any Cas9 protein, such as any one of the Cas9 amino acid sequences as provided herein, is a D.

[00452] A number of Cas9 sequences from various species were aligned to determine whether corresponding homologous amino acid residues of D10 and H840 of SEQ ID NO: 6 or SEQ ID NO: 567 can be identified in other Cas9 proteins, allowing the generation of Cas9 variants with corresponding mutations of the homologous amino acid residues. The alignment was carried out using the NCBI Constraint-based Multiple Alignment Tool (COBALT(accessible at st-va.ncbi.nlm.nih.gov/tools/cobalt), with the following parameters. Alignment parameters: Gap penalties -11,-1; End-Gap penalties -5,-1. CDD Parameters: Use RPS BLAST on; Blast E-value 0.003; Find Conserved columns and Recompute on. Query Clustering Parameters: Use query clusters on; Word Size 4; Max cluster distance 0.8; Alphabet Regular.

[00453] An exemplary alignment of four Cas9 sequences is provided below. The Cas9 sequences in the alignment are: Sequence 1 (SI): SEQ ID NO: 567 IWP_010922251I gi 499224711 Itype II CRISPR RNA-guided endonuclease Cas9 [Streptococcus pyogenes]; Sequence 2 (S2): SEQ ID NO: 568 IWP_039695303 Igi 746743737 Itype II CRISPR RNA-guided endonuclease Cas9 [Streptococcus gallolyticus]; Sequence 3 (S3): SEQ ID NO: 569 I WP_045635197 Igi 782887988 Itype II CRISPR RNA-guided endonuclease Cas9 [Streptococcus mitis]; Sequence 4 (S4): SEQ ID NO: 570 I5AXW_A Igi 924443546 I *Staphylococcus aureus* Cas9. The HNH domain (bold and underlined) and the RuvC domain (boxed) are identified for each of the four sequences. Amino acid residues 10 and 840 in SI and the homologous amino acids in the aligned sequences are identified with an asterisk following the respective amino acid residue.

```
51 1    – MDKK-YSIGLD*IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI– GALLFDSG––ETAEATRLKRTARRRYT  73
52 1    – MTKKNYSIGLD*IGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLL– GALLFDSG––ETAEATRLKRTARRRYT  74
53 1    – M-KKGYSIGLD*IGTNSVGFAVITDDYKVPSKKMKVLGNTDKRFIKKNLI– GALLFDEG––TTAEARRLKRTARRRYT  73
54 1    GSHMKRNYILGLD*IGITSVGYGII– DYET––––––––––––––––RDVIDAGVRLFKEANVENNEGRRSKRGARRLKR  61

51 74   RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL  153
52 75   RRKNRLRYLQEIFANEIAKVDESFFQRLDESFLTDDDKTFDSHPIFGNKAEEDAYHQKFPTIYHLRKHLADSSEKADLRL  154
53 74   RRKNRLRYLQEIFSEEMSKVDSSFFHRLDDSFLIPEDKRESKYPIFATLTEEKEYHKQFPTIYHLRKQLADSKEKTDLRL  153
54 62   RRRHRIQRVKKLL––––––––––––––FDYNLLTD––––––––––––––––––––HSELSGINPYEARVKGLSQKLSEEE  107
```

```
SI 154  IYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK   233
S2 155  VYLALAHMIKFRGHFLIEGELNAENTDVQKIFADFVGVYNRTFDDSHLSEITVDVASILTEKISKSRRLENLIKYYPTEK   234
S3 154  IYLALAHMIKYRGHFLYEEAFDIKNNDIQKIFNEFISIYDNTFEGSSLSGQNAQVEAIFTDKISKSAKRERVLKLFPDEK   233
S4 108  FSAALLHLAKRRG----------VHNVNEVEEDT----------                                    131

SI 234  KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT   313
S2 235  KNTLFGNLIALALGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSAKNLYDAILLSGILTVDDNST   314
S3 234  STGLFSEFLKLIVGNQADFKKHFDLEDKAPLQFSKDTYDEDLENLLGQIGDDFTDLFVSAKKLYDAILLSGILTVTDPST   313
S4 132  -----GNELS----------TKEQISRN----------                                           144

SI 314  KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM  — DGTEELLV   391
S2 315  KAPLSASMIKRYVEHHEDLEKLKEFIKANKSELYHDIFKDKNKNGYAGYIENGVKQDEFYKYLKNILSKIKIDGSDYFLD   394
S3 314  KAPLSASMIERYENHQNDLAALKQFIKNNLPEKYDEVFSDQSKDGYAGYIDGKTTQETFYKYIKNLLSKF  — EGTDYFLD   391
S4 145  ----SKALEEKYVAELQ----------LERLKKDG------                                        165

SI 392  KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE   471
S2 395  KIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKEKQDRIEKILTFRIPYYVGPLVRKDSRFAWAEYRSDE   474
S3 392  KIEREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEYYPFLKDNKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDE   471
S4 166  — EVRGSINRFKTSD--------YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGP — GEGSPFGW------K   227

SI 472  TITPWNFEEWDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL    551
S2 475  KITPWNFDKVIDKEKSAEKFITRMTLNDLYLPEEKVLPKHSHVYETYAVYNELTKIKYVNEQGKE-SFFDSNMKQEIFDH   553
S3 472  AIRPWNFEEIVDKASSAEDFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVNQ   551
S4 228  DIKEW----------YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEK ---LEYYEKFQI IEN 289

SI 552  LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR ---FNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFED   628
S2 554  VFKENRKVTKEKLLNYLNKEFPEYRIKDLIGLDKENKSFNASLGTYHDLKKIL-DKAFLDDKVNEEVIEDI  IKTLTLFED   632
S3 552  LFKENRKVTEKDI IHYLHN-VDGYDGIELKGIEKQ ---FNASLSTYHDLLKI IKDKEFMDDAKNEAILENIVHTLTIFED   627
S4 290  VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEF ---TNLKVYHD IKDITARKE11---ENAELLDQIAKILTIYQS   363

SI 629  REMIEERLKTYAHLFDDKVMKQLKR-RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED   707
S2 633  KDMIHERLQKYSDIFTANQLKKLER-RHYTGWGRLSYKLINGIRNKENNKTILDYLIDDGSANRNFMQLINDDTLPFKQI   711
S3 628  REMIKQRLAQYDSLFDEKVIKALTR-RHYTGWGKLSAKLINGICDKQTGNTILDYLIDDGKINRNFMQLINDDGLSFKEI   706
S4 364  SEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE ------LWHTNDNQIAIFNRLKLVP---------  428

51 708  IQKAQVSGQG|DSLHEHIANLAGSPAIKKGILQTVKWDELVKVMGRHKPENIVIEMA^RENQTT ------QKGQKNSRERM  781
52 712  IQKSQWGDV|DDIEAWHDLPGSPAIKKGILQSVKIVDELVKVMG-GNPDNIVIEMA|RENQTT ------NRGRSQSQQRL  784
53 707  IQKAQVIGKT|DDVKQWQELSGSPAIKKGILQSIKIVDELVKVMG-HAPESIVIEMA^RENQTT ------ARGKKNSQQRY  779
54 429  -KKVDLSQQK |EIPTTLVDDFILSPWKRSFIQSIKVINAIIKKYG — LPNDIIIELAREKNSKDAQKMINEMQKRNRQTN  505

51 782  KRIEEGIKELGSQIL-------KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD----YDVDH*IVPQSFLKDD  850
52 785  KKLQNSLKELGSNILNEEKPSYIEDKVENSHLQNDQLFLYYIQNGKDMYTGDELDIDHLSD----YDIDH*IIPQAFIKDD  860
53 780  KRIEDSLKILASGL---DSNILKENPTDNNQLQNDRLFLYYLQNGKDMYTGEALDINQLSS----YDIDH*IIPQAFIKDD  852
S4 506  ERIEEIIRTTGK----------ENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDH*IIPRSVSFDN  570

SI 851  SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN-LTKAERG|GL-SELD------KAGFIKRQLvj 922
```

```
52 861  SIDNRVLTSSAKNRGKSDDVPSLDIVRARK%EWVRLYKSGLISKRKFDN  -LTKAERG GL-TEAD-------KAGFIKRQLv] 932
53 853  SLDNRVLTSSKDNRGKSDNVPSIEWQKRK%FWQQLLDSKLISERKFNN  -LTKAERG GL-DERD-------KVGFIKRQLVI 924
54 571  SFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKE  YLLEERDINRFSVQKDFINRNLV| 650

51 923  |ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAWGTALIKKYP| 1002
52 933  |ETRQITKHVAQILDARFNTEHDENDKVIRDVKVITLKSNLVSQFRKDFEFYKVREINDYHHAHDAYLNAWGTALLKKYP| 1012
53 925  |ETRQITKHVAQILDARYNTEVNEKDKKNRTVKIITLKSNLVSNFRKEFRLYKVREINDYHHAHDAYLNAWAKAILKKYP| 1004
S4 651  IDTRYATRGLMNLLRSYFRVN- --------NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIA---------- 712

51 1003 KLESEFVYGDYKVYDVRKMIAKSEQ--EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG--- 1077
52 1013 KLASEFVYGEYKKYDIRKFITNSSD-----KATAKYFFYSNLMNFFKTKVKYADGTVFERPIIETNAD-GEIAWNKQ--- 1083
53 1005 KLEPEFVYGEYQKYDLKRYISRSKDPKEVEKATEKYFFYSNLLNFFKEEVHYADGTIVKRENIEYSKDTGEIAWNKE--- 1081
S4 713  NADFIFKEWKKLDKAKKVMENQM-------------------------FEEKQAESMPEIETEQEYKEIFITPHQIKI 764

51 1078 -----RDFATVRKVLSMPQVNIVKKTEVQT|GGFSKESILPKRNSDKLIARKKD---WDPKKYGGFDSPTVAYSVLWAKV 1149
52 1084 -----IDFEKVRKVLSYPQVNIVKKVETQT|GGFSKESILPKGDSDKLIPRKTKKVYWDTKKYGGFDSPTVAYSVFWADV 1158
53 1082 -----KDFAIIKKVLSLPQVNIVKKREVQT|GGFSKESILPKGNSDKLIPRKTKDILLDTTKYGGFDSPVIAYSILLIADI 1156
54 765  |HIKDFKDYKYSHRVDKKPNRELINDTLYST|RKDDKGNTLIVNNLNGLYDKDNDKL----KKLIN-KSP ----EKLLMYHH 835

51 1150 EKGKSKKLKSVKELLGI TIMERSSFEKNP I-DFLEAKG -----YKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKG 1223
52 1159 EKGKAKKLKTVKELVGI SIMERSFFEENPV-EFLENKG -----YHNIREDKLIKLPKYSLFEFEGGRRRLLASASELQKG 1232
53 1157 EKGKAKKLKTVKTLVGI TIMEKAAFEENP I-TFLENKG -----YHNVRKENILCLPKYSLFELENGRRRLLASAKELQKG 1230
54 836  DPQTYQKLK ---------LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKV 907

51 1224 NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKH ------ 1297
52 1233 NEMVLPGYLVELLYHAHRADNF -----NSTEYLNYVSEHKKEFEKVLSCVEDFANLYVDVEKNLSKIRAVADSM ------ 1301
53 1231 NEIVLPVYLTTLLYHSKNVHKL -----DEPGHLEYIQKHRNEFKDLLNLVSEFSQKYVLADANLEKIKSLYADN ------ 1299
54 908  VKLSLKPYRFD-VYLDNGVYKFV -----TVKNLDVIK — KENYYEVNSKAYEEAKKLKKISNQAEFIASFYNNDLIKING 979

51 1298 RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT ---------GLYETRI ----DLSQL 1365
52 1302 DNFSIEEISNSFINLLTLTALGAPADFNFLGEKIPRKRYTSTKECLNATLIHQSIT ---------GLYETRI ----DLSKL 1369
53 1300 EQADIEILANSFINLLTFTALGAPAAFKFFGKDIDRKRYTTVSEILNATLIHQSIT ---------GLYETWI ----DLSKL 1367
54 980  ELYRVI GVNNDLLNRIEVNMI DITYR-EYLENMNDKRPPRI IKT IASKT ---QSIKKYSTDILGNLYEVKSKKHPQI IKK 1055

SI 1366 GGD  1368
S2 1370 GEE  1372
S3 1368 GED  1370
S4 1056 G—   1056
```

[00454] The alignment demonstrates that amino acid sequences and amino acid residues that are homologous to a reference Cas9 amino acid sequence or amino acid residue can be identified across Cas9 sequence variants, including, but not limited to Cas9 sequences from different species, by identifying the amino acid sequence or residue that aligns with the reference

sequence or the reference residue using alignment programs and algorithms known in the art. This disclosure provides Cas9 variants in which one or more of the amino acid residues identified by an asterisk in SEQ ID NOs: 567-570 *(e.g.,* SI, S2, S3, and S4, respectively) are mutated as described herein. The residues D10 and H840 in Cas9 of SEQ ID NO: 6 that correspond to the residues identified in SEQ ID NOs: 567-570 by an asterisk are referred to herein as "homologous" or "corresponding" residues. Such homologous residues can be identified by sequence alignment, *e.g.,* as described above, and by identifying the sequence or residue that aligns with the reference sequence or residue. Similarly, mutations in Cas9 sequences that correspond to mutations identified in SEQ ID NO: 6 herein, *e.g.,* mutations of residues 10, and 840 in SEQ ID NO: 6, are referred to herein as "homologous" or "corresponding" mutations. For example, the mutations corresponding to the DIOA mutation in SEQ ID NO: 6 or S1 (SEQ ID NO: 567) for the four aligned sequences above are Dl 1A for S2, DIOA for S3, and D13A for S4; the corresponding mutations for H840A in SEQ ID NO: 6 or SI (SEQ ID NO: 567) are H850A for S2, H842A for S3, and H560A for S4.

EXAMPLE 6: Next Generation C to T Editors

[00455] Other familes of cytidine deaminases as alterantives to base etitor 3 (BE3) constructs were examined. The different C to T editors were developed to have a narrow or different editing window, alternate sequence specificity to expand targetable substrates, and to have higher activity.

[00456] Using the methods described in Example 4, the pmCDAl (cytidine deaminase 1 from *Petromyzon marinus)* activity at the HeK-3 site is evaluated (Figure 42). The pmCDAl -nCas9-UGI-NLS (nCas9 indicates the Cas9 nickase described herein) construct is active on some sites (e.g., the C bases on the complementary strand at position 9, 5, 4, and 3) that are not accessible with rAPOBECl (BE3).

[00457] The pmCDAl activity at the HeK-2 site is given in Figure 43. The pmCDAl -XTEN-nCas9-UGI-NLS construct is active on sites adjacent to "G," while rAPOBECl analog (BE3 construct) has low activity on "C"s that are adjacent to "G"s, e.g., the C base at position 11 on the complementary strand.

[00458] The percent of total sequencing reads with target C converted to T (Figure 44), C converted to A (Figure 45), and C converted to G (Figure 46) are shown for CDA and APOBECl (the BE3 construct).

[00459] The huAPOBEC3G activity at the HeK-2 site is shown in Figure 47. Two constructs were used: huAPOBEC3G-XTEN-nCas9-UGI-NLS and huAPOBEC3G*(D316R_D317R)-XTEN-nCas9-UGI-NLS. The huAPOBEC3G-XTEN-nCas9-UGI-NLS construct has different sequence specificity than rAPOBECl (BE3), as shown in Figure 47, the editing window appears narrow, as indicated by APOBEC3G's descreased activity at position 4 compared to APOBECl. Mutations made in huAPOBEC3G (D316R and D317R) increased ssDNA binding and resulted in an observable effect on expanding the sites which were edited (compare APOBEC3G with APOBEC3G_RR in Figure 47). Mutations were chosen based on APOBEC3G crystal structure, *see:* Holden *et al.,* Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implication. *Nature.* (2008); 121-4, the entire contents of which are incorporated herein by reference.

EXAMPLE 7: pmCDAl/huAPOBEC3G/rAPOBECl work in E. coli

[00460] LacZ selection optimization for the A to I conversion was performed using a bacterial strain with lacZ encoded on the F plasmid. A critical glutamic acid residue was mutated (e.g., GAG to GGG, Glu to Gly mutation) so that G to A by a cytidine deaminase would restore lacZ activity (Figure 48). Strain CC102 was selected for the selection assay. APOBECl and CDA constructs were used in a selection assay to optimize G to A conversion.

[00461] To evaluate the the effect of copy number of the plasmids encoding the deaminase constructs on lacZ reversion frequency, the CDA and APOBECl deaminases were cloned into 4 plasmids with different replication origins (hence different copy numbers), SC101, CloDF3, RSF1030, and PUC (copy number: PUC>RSF1030>CloDF3>SC101) and placed under an inducible promoter. The plasmids were individually transformed into E. coli cells harboring F plasmid containing the mutated LacZ gene. The expression of the deaminases were induced and LacZ activity was detected for each construct (Figure 49). As shown in Figure 49, CDA exhibited significantly higher activity than APOBECl in all instances, regardless of the plasmid copy number the deaminases were cloned in. Further, In terms of the copy number, the

deaminase activity was positively correlated with the copy number of the plasmid they are cloned in, i.e., PUC> CloDF3>SC101.

[00462] LacZ reversions were confirmed by sequencing of the genomic DNA at the lacZ locus. To obtain the genomic DNA containing the corrected LacZ gene, cells were grown media containg X-gal, where cells having LacZ activity form blue colonies. Blue colonies were selected and grown in minimial media containing lactose. The cells were spun down, washed, and re-plated on minimal media plates (lactose). The blue colony at the highest dilution was then selected, and its genomic DNA was sequenced at the lacZ locus (Figure 50).

[00463] A chloramphenicol reversion assay was designed to test the activity of different cytidine deaminases (e.g., CDA, and APOBEC1). A plasmid harboring a mutant CATl gene which confers chloramphenicol resistance to bacteria is constructed with RSF1030 as the replication origin. The mutant CATl gene encodings a CATl protein that has a H195R (CAC to CGC) mutation, rendering the protein inactive (Figure 51). Deamination of the C base-paired to the G base in the CGC codon would convert the codon back to a CAC codon, restoring the activity of the protein. As shown in Figure 52, CDA outperforms rAPOBEC in *E. coli* in restoring the acitivyt of the chloramphenicol reisitance gene. The minimum inhibitory concentration (MIC) of chlor in S1030 with the selection plasmid (pNMG_ch_5) was approximately 1 μg/mL. Both rAPOBEC-XTEN-dCas9-UGI and CDA-XTEN-dCas9-UGI induced DNA correction on the selection plasmid (Figure 53).

[00464] Next, the huAPOBEC3G-XTEN-dCas9-UGI protein was tested in the same assay. Interestingly, huAPOBEC3G-XTEN-dCas9-UGI exhibited different sequence specificity than the rAPOBEC l-XTEN-dCas9-UGI fusion protein. Only position 8 was edited with APOBEC3G-XTEN-dCas9-UGI fusion, as compared to the rAPOBEC H-XTEN-dCas9-UGIfusion (in which positions 3, 6, and 8 were edited) (Figure 54).

EXAMPLE 8: C to T Base Editors with Less Off Target Editing

[00465] Current base editing technologies allow for the sequence-specific conversion of a C:G base pair into a T:A base pair in genomic DNA. This is done via the direct catalytic conversion of cytosine to uracil by a cytidine deaminase enzyme and thus, unlike traditional genome editing technologies, does not introduce double-stranded DNA breaks (DSBs) into the DNA as a first step. See, Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016),

"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage." Nature *533,* 420-424; the entire contents of which are incorporated by reference herein. Instead, catalytically dead SpCas9 (dCas9) or a SpCas9 nickase (dCas9(A840H)) is tethered to a cytidine deaminase enzyme such as rAPOBECl, pmCDAl, or hAPOBEC3G. The genomic locus of interest is encoded by an sgRNA, and DNA binding and local denaturation is facilitated by the dCas9 portion of the fusion. However, just as wt dCas9 and wt Cas9 exhibit off-target DNA binding and cleavage, current base editors also exhibit C to T editing at Cas9 off-target loci, which limits their therapeutic usefulness.

[00466] It has been reported that the introduction of just three to four mutations into SpCas9 that neutralize nonspecific electrostatic interactions between the protein and the sugar-phosphate backbone of its target DNA, increases the DNA binding specificity of SpCas9. See, Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016) "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." Nature *529,* 490-495; and Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2015) "Rationally engineered Cas9 nucleases with improved specificity. Science *351,* 84-88; the entire contents of each are hereby incorporated by reference herein. Four reported neutralizing mutations were therefore incorporated into the initially reported base editor BE3 (SEQ ID NO: 48), and found that off-target C to T editing of this enzyme is also drastically reduced (Figure 55), with no decrease in on-target editing (Figure 56).

[00467] As shown in Figure 55, HEK293T cells were transfected with plasmids expressing BE3 or HF-BE3 and a sgRNA matching the EMX1 sequence using Lipofectamine 2000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target locus, plus the top ten known Cas9 off-target loci for the EMX1 sgRNA, as previously determined by Joung and coworkers using the GUIDE-seq method. See Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C , Iafrate, A J., Le, L.P., *et al.* (2015) "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases." *Nat Biotech 33,* 187-197; the entire contents of which are incorporated by reference herein. EMX1 off-target 5 locus did not amplify and is not shown. Sequences of the on-target and off-target protospacers and protospacer adjacent motifs (PAMs) are displayed (Figure 55). Cellular C to T conversion percentages,

defined as the percentage of total DNA sequencing reads with T at each position of an original C within the protospacer, are shown for BE3 and HF-BE3.

[00468] In Figure 56, HEK293T cells were transfected with plasmids expressing BE3 or HF-BE3 and sgRNAs matching the genomic loci indicated using Lipofectamine 2000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target loci. The percentage of total DNA sequencing reads with all four bases at the target Cs within each protospacer are shown for treatment with BE3 or HF-BE3 (Figure 56). Frequencies of indel formation are shown as well.

[00469] Primary Protein Sequence of HF-BE3 (SEQ ID NO: 48):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
ALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD
ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

EXAMPLE 9: Development of Base Editors that Use Cas9 Variants and Modulation of the Base Editor Processivity to Increase the Target Range and Precision of the Base Editing Technology

[00470] Unlike traditional genome editing platforms, base editing technology allows precise single nucleotide changes in the DNA without inducing double-stranded breaks(DSBs). See, Komor, A. C. *et al. Nature* **533,** 420-424 (2016). The current generation of base editor uses the NGG PAM exclusively. This limits its ability to edit desired bases within the genome, as the base editor needs to be placed at a precise location where the target base is placed within a 4-

base region (the 'deamination window'), approximately 15 bases upstream of the PAM . See, Komor, A. C. *et al. Nature* **533,** 420-424 (2016). Moreover, due to the high processivity of cytidine deaminase, the base editor may convert all cytidines within its deamination window into thymidines, which could induce amino acid changes other than the one desired by the researcher. See, Komor, A. C. *et al. Nature* **533,** 420-424 (2016).

Expanding the scope of base editing through the development of base editors with Cas9 variants

[00471] Cas9 homologs and other RNA-guided DNA binders that have different PAM specificities were incorporated into the base editor architecture. See, Kleinstiver, B. P. *et al. Nature* **523,** 481-485 (2015); Kleinstiver, B. P. *et al. Nature Biotechnology* **33,** 1293-1298 (2015); and Zetsche, B. *et al. Cell* **163,** 759-771 (2015); the entire contents of each are incorporated by reference herein. Furthermore, innovations that have broadened the PAM specificities of various Cas9 proteins were also incorporated to expand the target reach of the base editor even more. See, Kleinstiver, B. P. *et al. Nature* **523,** 481-485 (2015); and Kleinstiver, B. P. *et al. Nature Biotechnology* **33,** 1293-1298 (2015). The current palette of base editors is summarized in Table 4.

Table 4. New base editors made from Cas9 Variants

Species	PAM	Base Editor Name	Reference for Cas9 variant
S. pyogenes	...NGG	BE3	Wild-type
	...NGA	VQR BE3 or EQR BE3	Kleinstiver, B. P. *et al.*
	...NGCG	VRER BE3	Kleinstiver, B. P. *et al.*
S. aureus	...NNGRRT	SaBE3	Wild-type
	...NNNRRT	SaKKH BE3	Kleinstiver, B. P. *et al.*
L. bacterium	TTTN...	dCpf1 BE2	Zetsche, B. *et al.*

Modulating base editor's processivity through site-directed mutagenesis of rAPOBECl

[00472] It was reasoned that the processivity of the base editor could be modulated by making point mutations in the deaminase enzyme. The incorporatation of mutations that slightly reduce the catalytic activity of deaminase in which the base editor could still catalyze on average one round of cytidine deamination but was unlikely to access and catalyze another deamination

within the relevant timescale were pursued. In effect, the resulting base editor would have a narrower deamination window.

[00473] rAPOBEC1 mutations probed in this work are listed in Table 5. Some of the mutations resulted in slight apparent impairment of rAPOBEC1 catalysis, which manifested as preferential editing of one cytidine over another when multiple cytidines are found within the deamination window. Combining some of these mutations had an additive effect, allowing the base editor to discriminate substrate cytidines with higher stringency. Some of the double mutants and the triple mutant allowed selective editing of one cytidine among multiple cytidines that are right next to one another (Figure 57).

Table 5. rAPOBEC1 Point Mutations Investigated

rAPOBEC1 mutation studied in this work	Corresponding mutation in APOBEC3G	Reference
W122R / W122R	D315R / D315R	Holden, L. G. et al.
R126A	R320A	Chen, K-M. et al.
R126E	R320E	Chen, K-M. et al.
R118A	R313A	Chen, K-M. et al.
W90A	W285A	Chen, K-M. et al.
W90Y	W285Y	
R132E	R326E	

Base Editor PAM Expansion and Processivity Modulation

[00474] The next generation of base editors were designed to expand editable cytidines in the genome by using other RNA-guided DNA binders (Figure 58). Using a NGG PAM only allows for a single target within the "window" whereas the use of multiple different PAMs allows for Cas9 to be positioned anywhere to effect selective deamination. A variety of new base editors have been created from Cas9 variants (Figure 59 and Table 4). Different PAM sites (NGA, Figure 60; NGCG, Figure 61; NNGRRT, Figure 62; and NNHRRT, Figure 63) were explored. Selective deamination was successfully achieved through kinetic modulation of cytidine deaminase point mutagenesis (Figure 65 and Table 5).

[00475] The effect of various mutations on the deamination window was then investigated in cell culture using spacers with multiple cytidines (Figures 66 and 67).

[00476] Further, the effect of various mutations on different genomic sites with limited numbers of cytidines was examined (Figures 68 to 71). It was found that approximately one cytidine will be edited within the deamination windown in the spacer, while the rest of the cytidines will be left intact. Overall, the preference for editing is as follows: $C_6 > C_5 \gg C_7 \approx C_4$.

Base Editing Using Cpfl

[00477] Cpfl, a Cas9 homolog, can be obtained as AsCpfl, LbCpfl, or from any other species. Schematics of fusion constructs, including BE2 and BE3 equivalents, are shown in Figure 73. The BE2 equivalent uses catalytically inactive Cpf2 enzyme (dCpfl) instead of Cas9, while the BE3 equivalent includes the Cpfl mutant, which nicks the target strand. The bottom schematic depicts different fusion architectures to combine the two innovations illustrated above it (Figure 73). The base editing results of HEK293T cell TTTN PAM sites using Cpfl BE2 were examined with different spacers (Figures 64A to 64C). In some embodiments, Cpfl may be used in place of a Cas9 domain in any of the base editors provided herein. In some embodiments, the Cpfl is a protein that is at lesst 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, or 99.5% identical to SEQ ID NO 9.

Full Protein Sequence of Cpfl (SEQ ID NO: 9):

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISED
LLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQS
KDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYE
SLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNG
ENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKT
VEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKE
QELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQG
KKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKI
RNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGE
GYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYK
QSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIK
DKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNI
IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFK
RGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTS
KICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEL
DYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQ
NRNN

EXAMPLE 10: Increased Fidelity of Base Editing

[00478] Examining the difference between plasmid delivery of BE3 and HF-BE3, it was found that the two edit on-target loci with comparable efficiency (Figures 74 and 75). However, HF-BE3 edited off-target loci much less than BE3, meaning that HF-BE3 has a much higher DNA specificity than BE3 (Figure 76). Deaminase protein lipofection to HEK cells demonstrated that protein delivery of BE3 results in comparable on-target activity, but much better specificity, than plasmid DNA delivery of BE3. Using improved transfection procedures and better plasmids (n=2), the experiment used the following conditions: protein delivery was 125 nM Cas9:sgRNA complex, plasmid delivery was 750ng BE3/HF-BE3 plasmid + 250ng sgRNA plasmid, and lipofection was with 1.5μL of Lipofectamine 2000 per well. EMX-1 off target site 2 and FANCF off-target site 1 showed the most off-target editing with BE3, compared to all of the off-targets assayed (Figures 77 and 78), while HEK-3 showed no significant editing at off-targets for any of the delivery methods (Figure 79). HEK-4 shows some C-to-G editing on at the on-target site, while its off-target sites 1, 3, and 4 showed the most off-target editing of all the assayed sites (Figure 80).

Delivery of BE3 Protein via Micro-injection to Zebrafish

[00479] TYR guide RNAs were tested in an *in vitro* assay for sgRNA activity (Figures 81 and 82). The % HTS reads shows how many C residues were converted to T residues during a 2h incubation with purified BE3 protein and PCR of the resulting product. Experiments used an 80-mer synthetic DNA substate with the target deamination site in 60bp of its genomic context. This is not the same as % edited DNA strands because only one strand was nicked, so the product is not amplified by PCR. The proportion of HTS reads edited is equal to x/(2-x), where x is the actual proportion of THS reads edited. For 60% editing, the actual proportion of bases edited is 75%. "Off target" is represents BE3 incubated with the same DNA substrate, while bound to an off-target sgRNA. It was found sgRNAs sgRH_13, sgHR_17, and possibly sgHR_16 appeared to be promising targets for *in vivo* injection experiments.

[00480] The delivery of BE3 protein in was tested *in vivo* in zebrafish. Zebrafish embryos (n=16-24) were injected with either scramled sgRNA, sgHR_13, sgHR_16, or sgHR_17 and purified BE3. Three embryos from each condition were analyzed independently (single embryo) and for each condition, all of the injected embryos were pooled and sequenced as a pool. The results are shown in Figures 83 to 85.

EXAMPLE 11: Uses of Base Editors to Treat Disease

[00481] Base editors or complexes provided herein *(e.g.,* BE3) may be used to modify nucleic acids. For example, base editors may be used to change a cytosine to a thymine in a nucleic acid *(e.g.,* DNA). Such changes may be made to, *inter alia,* alter the amino acid sequence of a protein, to destroy or create a start codon, to create a stop codon, to distupt splicing donors, to disrupt splicing acceptors or edit regulatory sequences. Examples of possible nucleotide changes are shown in Figure 86.

[00482] Base editors or complexes provided herein *(e.g.,* BE3) may be used to edit an isoform of Apolipoprotein E in a subject. For example, an Apolipoprotein E isoform may be edited to yield an isoform associated with a lower risk of developing Alzheimer's disease. Apolipoprotein E has four isoforms that differ at amino acids 112 and 158. APOE4 is the largest and most common genetic risk factor for late-onset Alzheimer's disease. Arginine residue 158 of APOE4, encoded by the nucleic acid sequence CGC, may be changed to a cysteine by using a base editor *(e.g.,* BE3) to change the CGC nucleic acid sequence to TGC, which encodes cysteine at residue 158. This change yields an APOE3r isoform, which is associated with lower Alzheimer's disease risk. See Figure 87.

[00483] It was tested whether base editor BE3 could be used to edit APOE4 to APOE3r in mouse astrocytes (Figure 88). APOE 4 mouse astrocytes were nucleofected with Cas9 + template or BE3, targeting the nucleic acid encoding Arginine 158 of APOE4. The Cas9 + template yielded only 0.3% editing with 26% indels, while BE3 yielded 75% editing with 5% indels. Two additional base-edited cytosines are silent and do not yield changes to the amino acid sequence (Figure 88).

[00484] Base editors or complexes provided herein may be used to treat prion protein diseases such as Creutzfeldt-Jakob disease and fatal familial insomnia, for example, by introducing mutations into a PRNP gene. Reverting PRNP mutations may not yield therapeutic results, and intels in PRNP may be pathogenic. Accordingly, it was tested whether PRNP could be mutated using base editors *(e.g.,* BE3) to introduce a premature stop codon in the PRNP gene. BE3, associated with its guide RNA,was introduced into HEK cells or glioblastoma cells and was capable of editing the PRNP gene to change the encoded arginine at residue 37 to a stop codon. BE3 yielded 41% editing (Figure 89).

[00485] Additional genes that may be edited include the following: *APOE* editing of Arg 112 and Arg 158 to treat increased Alzheimer's risk; *APP* editing of Ala 673 to decrease Alzheimer's risk; *PRNP* editing of Arg 37 to treat fatal familial insomnia and other prion protein diseases; *DMD* editing of the exons 23 and 51 splice sites to treat Duchenne muscular dystrophy; *FTO* editing of intron 1 to treat obesity risk; *PDS* editing of exon 8 to treat Pendred syndrome (genetic deafness); *TMC1* editing of exon 8 to treat congenital hearing loss; *CYBB* editing of various patient-relevant mutations to treat chronic granulomatous disease. Additional diseases that may be treated using the base editors provided herein are shown in Table 6, below.

[00486] UGI also plays a key role. Knocking out UDG (which UGI inhibits) was shown to dramatically improve the cleanliness and efficiency of C to T base editing (Figure 90). Furthermore, base editors with nickase and without UGI were shown to produce a mixture of outcomes, with very high indel rates (Figure 91).

EXAMPLE 12: Expanding the Targeting Scope of Base Editing

[00487] Base editing is a new approach to genome editing that uses a fusion protein containing a catalytically defective *Streptococcus pyogenes* Cas9, a cytidine deaminase, and an inhibitor of base excision repair to induce programmable, single-nucleotide C-^T (or G-^A) changes in DNA without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions [1]. The development of five new C-^T (or G-^A) base editors that use natural and engineered Cas9 variants with different protospacer-adjacent motif (PAM) specificities to expand the number of sites that can be targeted by base editing by 2.5-fold are described herein. Additionally, new base editors containing mutated cytidine deaminase domains that narrow the width of the apparent editing window from approximately 5 nucleotides to 1 or 2 nucleotides were engineered, enabling the discrimination of neighboring C nucleotides that would previously be edited with comparable efficiency. Together, these developments substantially increase the targeting scope of base editing.

[00488] CRISPR-Cas9 nucleases have been widely used to mediate targeted genome editing[2]. In most genome editing applications, Cas9 forms a complex with a single guide RNA (sgRNA) and induces a double-stranded DNA break (DSB) at the target site specified by the sgRNA sequence. Cells primarily respond to this DSB through the non-homologuous end-joining (NHEJ) repair pathway, which results in stochastic insertions or deletions (indels) that can cause

frameshift mutations that disrupt the gene. In the presence of a donor DNA template with a high degree of homology to the sequences flanking the DSB, gene correction can be achieved through an alternative pathway known as homology directed repair (HDR).[3,4] Unfortunately, under most non-perturbative conditions HDR is inefficient, dependent on cell state and cell type, and dominated by a larger frequency of indels.[3,4] As most of the known genetic variations associated with human disease are point mutations[5], methods that can more efficiently and cleanly make precise point mutations are needed.

[00489] Base editing, which enables targeted replacement of a C:G base pair with a T:A base pair in a programmable manner without inducing DSBs[1], has been recently described. Base editing uses a fusion protein between a catalytically inactivated (dCas9) or nickase form of *Streptococcus pyogenes* Cas9 (SpCas9), a cytidine deaminase such as APOBEC1, and an inhibitor of base excision repair such as uracil glycosylase inhibitor (UGI) to convert cytidines into uridines within a five-nucleotide window specified by the sgRNA.[1] The third-generation base editor, BE3, converts C:G base pairs to T:A base pairs, including disease-relevant point mutations, in a variety of cell lines with higher efficiency and lower indel frequency than what can be achieved using other genome editing methods[1]. Subsequent studies have validated the deaminase-dCas9 fusion approach in a variety of settings[6,7].

[00490] Efficient editing by BE3 requires the presence of an NGG PAM that places the target C within a five-nucleotide window near the PAM-distal end of the protospacer (positions 4-8, counting the PAM as positions 21-23)[1]. This PAM requirement substantially limits the number of sites in the human genome that can be efficiently targeted by BE3, as many sites of interest lack an NGG 13- to 17- nucleotides downstream of the target C. Moreover, the high activity and processivity of BE3 results in conversion of all Cs within the editing window to Ts, which can potentially introduce undesired changes to the target locus. Herein, new C:G to T:A base editors that address both of these limitations are described.

[00491] It was thought that any Cas9 homolog that binds DNA and forms an "R-loop" complex[8] containing a single-stranded DNA bubble could in principle be converted into a base editor. These new base editors would expand the number of targetable loci by allowing non-NGG PAM sites to be edited. The Cas9 homolog from *Staphylococcus aureus* (SaCas9) is considerably smaller than SpCas9 (1053 vs. 1368 residues), can mediate efficient genome editing in mammalian cells, and requires an NNGRRT PAM[9]. SpCas9 was replaced with SaCas9 in BE3

to generate SaBE3 and transfected HEK293T cells with plasmids encoding SaBE3 and sgRNAs targeting six human genomic loci (Figures 92A and 92B). After 3 d, the genomic loci were subjected to high-throughput DNA sequencing (HTS) to quantify base editing efficiency. SaBE3 enabled C to T base editing of target Cs at a variety of genomic sites in human cells, with very high conversion efficiencies (approximately 50-75% of total DNA sequences converted from C to T, without enrichment for transfected cells) arising from targeting Cs at positions 6-11. The efficiency of SaBE3 on NNGRRT-containing target sites in general exceeded that of BE3 on NGG-containing target sites[1]. Perhaps due to its higher average efficiency, SaBE3 can also result in detectable base editing at target Cs at positions outside of the canonical BE3 activity window (Figure 92C). In comparison, BE3 showed significantly reduced editing under the same conditions (0-11%), in accordance with the known SpCas9 PAM preference (Figure 106A)[10]. These data show that SaBE3 can facilitate very efficient base editing at sites not accessible to BE3.

[00492] The targeting range of base editors was further expanded by applying recently engineered Cas9 variants that expand or alter PAM specificities. Joung and coworkers recently reported three SpCas9 mutants that accept NGA (VQR-Cas9), NGAG (EQR-Cas9), or NGCG(VRER-Cas9) PAM sequences[11]. In addition, Joung and coworkers engineered a SaCas9 variant containing three mutations (SaKKH-Cas9) that relax its PAM requirement to NNNRRT[12]. The SpCas9 portion of BE3 was replaced with these four Cas9 variants to produce VQR-BE3, EQR-BE3, VRER-BE3, and SaKKH-BE3, which target NNNRRT, NGA, NGAG, and NGCG PAMs respectively. HEK293T cells were transfected with plasmids encoding these constructs and sgRNAs targeting six genomic loci for each new base editor, and measured C to T base conversions using HTS.

[00493] SaKKH-BE3 edited sites with NNNRRT PAMs with efficiencies up to 62% of treated, non-enriched cells (Figure 92D). As expected, SaBE3 was unable to efficiently edit targets containing PAMs that were NNNHRRT (where H = A, C, or T) (Figure 92D). VQR-BE3, EQR-BE3, and VRER-BE3 exhibited more modest, but still substantial base editing efficiencies of up to 50% of treated, non-enriched cells at genomic loci with the expected PAM requirements with an editing window similar to that of BE3 (Figures 92E and 92F). Base editing efficiencies of VQR-BE3, EQR-BE3, and VRER-BE3 in general closely paralleled the reported PAM requirements of the corresponding Cas9 nucleases; for example, EQR-BE3 was unable to

efficiently edit targets containing NGAH PAM sequences (Figure 92F). In contrast, BE3 was unable to edit sites with NGA or NGCG PAMs efficiently (0-3%), likely due to its PAM restrictions (Figure 106B).

[00494] Collectively, the properties of SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, and VRER-BE3 establish that base editors exhibit a modularity that facilitates their ability to exploit Cas9 homologs and engineered variants.

[00495] Next, base editors with altered activity window widths were developed. All Cs within the activity window of BE3 can be efficiently converted to Ts[1]. The ability to modulate the width of this window would be useful in cases in which it is important to edit only a subset of Cs present in the BE3 activity window.

[00496] The length of the linker between APOBEC1 and dCas9 was previously observed to modulate the number of bases that are accessible by APOBEC1 *in vitro*[1]. In HEK293T cells, however, varying the linker length did not significantly modulate the width of the editing window, suggesting that in the complex cellular milieu, the relative orientation and flexibility of dCas9 and the cytidine deaminase are not strongly determined by linker length (Figure 96). Next, it was thought that truncating the 5' end of the sgRNA might narrow the base editing window by reducing the length of single-stranded DNA accessible to the deaminase upon formation of the RNA-DNA heteroduplex. HEK293T cells were co-transfected with plasmids encoding BE3 and sgRNAs of different spacer lengths targeting a locus with multiple Cs in the editing window. No consistent changes in the width of base editing when using truncated sgRNAs with 17- to 19-base spacers were observed (Figures 95A to 95C). Truncating the sgRNA spacer to fewer than 17 bases resulted in large losses in activity (Figure 95A).

[00497] As an alternative approach, it was thought that mutations to the deaminase domain might narrow the width of the editing window through multiple possible mechanisms. First, some mutations may alter substrate binding, the conformation of bound DNA, or substrate accessibility to the active site in ways that reduce tolerance for non-optimal presentation of a C to the deaminase active site. Second, because the high activity of APOBEC1 likely contributes to the deamination of multiple Cs per DNA binding event,[1,13,14] mutations that reduce the catalytic efficiency of the deaminase domain of a base editor might prevent it from catalyzing successive rounds of deamination before dissociating from the DNA. Once any C:G to T:A editing event has taken place, the sgRNA no longer perfectly matches the target DNA sequence

and re-binding of the base editor to the target locus should be less favorable. Both strategies were tested in an effort to discover new base editors that distinguish among multiple cytidines within the original editing window.

[00498] Given the absence of an available APOBEC1 structure, several mutations previously reported to modulate the catalytic activity of APOBEC3G, a cytidine deaminase from the same family that shares 42% sequence similarity of its active site-containing domain to that of APOBEC1, were identified[15]. Corresponding APOBEC1 mutations were incorporated into BE3 and evaluated their effect on base editing efficiency and editing window width in HEK293T cells at two C-rich genomic sites containing Cs at positions 3, 4, 5, 6, 8, 9, 10, 12, 13, and 14 (site A); or containing Cs at positions 5 ,6, 7, 8, 9, 10, 11, and 13 (site B).

[00499] The APOBEC 1 mutations R 118A and W90A each led to dramatic loss of base editing efficiency (Figure 97C). R132E led to a general decrease in editing efficiency but did not change the substantially narrow the shape of the editing window (Figure 97C). In contrast, several mutations that narrowed the width of the editing window while maintaining substantial editing efficiency were found (Figures 93A and 97C). The "editing window width" was defined to represent the artificially calculated window width within which editing efficiency exceeds the half-maximal value for that target. The editing window width of BE3 for the two C-rich genomic sites tested was 5.0 (site A) and 6.1 (site B) nucleotides.

[00500] R126 in APOBEC 1 is predicted to interact with the phosphate backbone of ssDNA[13]. Previous studies have shown that introducing the corresponding mutation into APOBEC3G decreased catalysis by at least 5-fold[14]. Interestingly, when introduced into APOBEC 1 in BE3, R126A and R126E increased or maintained activity relative to BE3 at the most strongly edited positions (C5, C6, and C7), while decreasing editing activity at other positions (Figures 93A and 97C). Each of these two mutations therefore narrowed the width of the editing window at site A and site B to 4.4 and 3.4 nucleotides (R126A), or to 4.2 and 3.1 nucleotides (R126E), respectively (Figures 93A and97C).

[00501] W90 in APOBEC 1 (corresponding to W285 in APOBEC3G) is predicted to form a hydrophobic pocket in the APOBEC3G active site and assist in substrate binding[13]. Mutating this residue to Ala abrogated APOBEC3G's catalytic activity[13]. In BE3, W90A almost completely abrogated base editing efficiency (Figure 97C). In contrast, it was found that W90Y only modestly decreased base editing activity while narrowing the editing window width at site A and

site B to 3.8 and 4.9 nucleotides, respectively (Figure 93A). These results demonstrate that mutations to the cytidine deaminase domain can narrow the activity window width of the corresponding base editors.

[00502] W90Y, R126E, and R132E, the three mutations that narrowed the editing window without drastically reducing base editing activity, were combined into doubly and triply mutated base editors. The double mutant W90Y+R126E resulted in a base editor (YE1-BE3) with BE3-like maximal editing efficiencies, but substantially narrowed editing window width (width at site A and site B = 2.9 and 3.0 nucleotides, respectively (Figure 93A). The W90Y+R132E base editor (YE2-BE3) exhibited modestly lower editing efficiencies (averaging 1.4-fold lower maximal editing yields across the five sites tested compared with BE3), and also substantially narrowed editing window width (width at site A and site B = 2.7 and 2.8 nucleotides, respectively) (Figure 97C). The R126E+R132E double mutant (EE-BE3) showed similar maximal editing efficiencies and editing window width as YE2-BE3 (Figure 97C). The triple mutant W90Y+R126E+R132E (YEE-BE3) exhibited 2.0-fold lower average maximal editing yields but very little editing beyond the C6 position and an editing window width of 2.1 and 1.4 nucleotides for site A and site B, respectively (Figure 97C). These data taken together indicate that mutations in the cytidine deaminase domain can strongly affect editing window widths, in some cases with minimal or only modest effects on editing efficiency.

[00503] The base editing outcomes of BE3, YE1-BE3, YE2-BE3, EE-BE3, and YEE-BE3 were further compared in HEK293T cells targeting four well-studied human genomic sites that contain multiple Cs within the BE3 activity window[1]. These target loci contained target Cs at positions 4 and 5 (HEK site 3), positions 4 and 6 (HEK site 2), positions 5 and 6 (EMXl), or positions 6, 7, 8, and 11 (FANCF). BE3 exhibited little (< 1.2-fold) preference for editing any Cs within the position 4-8 activity window. In contrast, YE1-BE3, exhibited a 1.3-fold preference for editing C5 over C4 (HEK site 3), 2.6-fold preference for C6 over C4 (HEK site 2), 2.0-fold preference for C5 over C6 (EMXl), and 1.5-fold preference for C6 over C7 (FANCF) (Figure 93B). YE2-BE3 and EE-BE3 exhibited somewhat greater positional specificity (narrower activity window) than YE1-BE3, averaging 2.4-fold preference for editing C5 over C4 (HEK site 3), 9.5-fold preference for C6 over C4 (HEK site 2), 2.9-fold preference for C5 over C6 (EMXl), and 2.6-fold preference for C7 over C6 (FANCF) (Figure 93B). YEE-BE3 showed the greatest positional selectivity, with a 2.9-fold preference for editing C5 over C4 (HEK site 3), 29.7-fold

preference for C6 over C4 (HEK site 2), 7.9-fold preference for C5 over C6 (EMX1), and 7.9-fold preference for C7 over C6 (FANCF) (Figure 93B). The findings establish that mutant base editors can discriminate between adjacent Cs, even when both nucleotides are within the BE3 editing window.

[00504] The product distributions of these four mutants and BE3 were further analyzed by HTS to evaluate their apparent processivity. BE3 generated predominantly T4-T5 (HEK site 3), T4-T6 (HEK site 2), and T5-T6 (EMX1) products in treated HEK293T cells, resulting in, on average, 7.4-fold more products containing two Ts, than products containing a single T. In contrast, YE1-BE3, YE2-BE3, EE-BE3, and YEE-BE3 showed substantially higher preferences for singly edited C4-T5, C4-T6, and T5-C6 products (Figure 93C). YE1-BE3 yielded products with an average single-T to double-T product ratio of 1.4. YE2-BE3 and EE-BE3 yielded products with an average single-T to double-T product ratio of 4.3 and 5.1, respectively (Figure 93C). Consistent with the above results, the YEE-BE3 triple mutant favored single-T products by an average of 14.3-fold across the three genomic loci. (Figure 93C). For the target site in which only one C is within the target window (HEK site 4, at position C5), all four mutants exhibited comparable editing efficiencies as BE3 (Figure 98). These findings indicate that these BE3 mutants have decreased apparent processivity and can favor the conversion of only a single C at target sites containing multiple Cs within the BE3 editing window. These data also suggest a positional preference of C5 > C6 > C7 ≈ C4 for these mutant base editors, although this preference could differ depending on the target sequence.

[00505] The window-modulating mutations in APOBEC1 were applied to VQR-BE3, allowing selective base editing of substrates at sites targeted by NGA PAM (Figure 107A). However, when these mutations were applied to SaKKH-BE3, a linear decrease in base editing efficiency was observed without the improvement in substrate selectivity, suggesting a different kinetic equilibrium and substrate accessibility of this base editor than those of BE3 and its variants (Figure 107B).

[00506] The five base editors with altered PAM specificities described in this study together increase the number of disease-associated mutations in the ClinVar database that can in principle be corrected by base editing by 2.5-fold (Figures 94A and 94B). Similarly, the development of base editors with narrowed editing windows approximately doubles the fraction of ClinVar entries with a properly positioned NGG PAM that can be corrected by base editing without

comparable modification of a non-target C (from 31% for BE3 to 59% for YEE-BE3) (Figures 94A and 94B).

[00507] In summary, the targeting scope of base editing was substantially expanded by developing base editors that use Cas9 variants with different PAM specificities, and by developing a collection of deaminase mutants with varying editing window widths. In theory, base editing should be possible using other programmable DNA-binding proteins (such as Cpfl [16]) that create a bubble of single-stranded DNA that can serve as a substrate for a single-strand-specific nucleotide deaminase enzyme.

Materials and Methods

[00508] *Cloning.* PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs). Plasmids for BE and sgRNA were constructed using USER cloning (New England Biolabs), obtained from previously reported plasmids [1]. DNA vector amplification was carried out using NEB lObeta competent cells (New England Biolabs).

[00509] *Cell culture.* HEK293T (ATCC CRL-3216) were cultured in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher) supplemented with 10% (v/v) fetal bovine serum (FBS), at 37 °C with 5% $C0_2$. Immortalized rat astrocytes containing the ApoE4 isoform of the *APOE* gene (Taconic Biosciences) were maintained in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS) and 200 μg/mL Geneticin (ThermoFisher Scientific).

[00510] *Transfections.* HEK293T cells were seeded on 48-well collagen-coated BioCoat plates (Corning) and transfected at approximately 85% confluency. 750 ng of BE and 250 ng of sgRNA expression plasmids were transfected using 1.5 μï of Lipofectamine 2000 (ThermoFisher Scientific) per well according to the manufacturer's protocol.

[00511] *High-throughput DNA sequencing of genomic DNA samples.* Transfected cells were harvested after 3 d and the genomic DNA was isolated using the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) according to the manufacturer's instructions. Genomic regions of interest were amplified by PCR with flanking HTS primer pairs listed in the Supplementary Sequences. PCR amplification was carried out with Phusion hot-start II DNA polymerase (ThermoFisher) according to the manufacturer's instructions. PCR products were purified using RapidTips (Diffinity Genomics). Secondary PCR was performed to attach sequencing adaptors. The products were gel-purified and quantified using the KAPA Library

Quantification Kit-Illumina (KAPA Biosystems). Samples were sequenced on an Illumina MiSeq as previously described [1].

[00512] *Data analysis.* Nucleotide frequencies were assessed using a previously described MATLAB script [1]. Briefly, the reads were aligned to the reference sequence via the Smith-Waterman algorithm. Base calls with Q-scores below 30 were replaced with a placeholder nucleotide (N). This quality threshold results in nucleotide frequencies with an expected theoretical error rate of 1 in 1000.

[00513] Analyses of base editing processivity were performed using a custom python script. This program trims sequencing reads to the 20 nucleotide protospacer sequence as determined by a perfect match for the 7 nucleotide sequences that should flank the target site. These targets were then consolidated and sorted by abundance to assess the frequency of base editing products.

[00514] Bioinformatic analysis of the ClinVar database of human disease-associated mutations was performed in a manner similar to that previously described but with small adjustments [1]. These adjustments enable the identification of targets with PAMs of customizable length and sequence. In addition, this improved script includes a priority ranking of target C positions (C5 > C6 > C7 >C8 ≈ C4), thus enabling the identification of target sites in which the on-target C is either the only cytosine within the window or is placed at a position with higher predicted editing efficiency than any off-target C within the editing window.

References for Example 12

1 Komor, A. C. *et al.* Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533,** 420-424 (2016).
2 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32,** 347-355 (2014).
3 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819-823 (2013).
4 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8,** 2281-2308 (2013).
5 Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44,** D862-D868 (2015).
6 Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353,** aaf8729-l-8 (2016).
7 Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* doi: 10.1038/nmeth. 4027 (2016).
8 Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351,** 867-71 (2016).

9 Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520,** 186-191 (2015).
10 Zhang, Y. *et al.* Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci. Rep.* **4,** (2014).
11 Kleinstiver, B. P. *et. al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523,** 481-485 (2015).
12 Kleinstiver, B. P. *et. al.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33,** 1293-1298 (2015).
13 Holden, L. G. et al. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **452,** 121-124 (2008).
14 Chen, K.-M. et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452,** 116-119 (2008).
15 Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA Editing Enzyme APOBECl and Some of Its Homologs Can Act as DNA Mutators. *Molecular Cell* **10,** 1247-1253 (2002).
16 Zetsche, B. *et al.* Cpfl Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163,** 759-771 (2015).

EXAMPLE 13

[00515] Using improved transfection procedures and better plasmids, biological replicates (n=3) were used to install the four HF mutations into the Cas9 portion of BE3. The muations do not significantly effect on-targeting editing with plasmid delivery (Figure 99). At the tested concentration, BE3 protein delivery works; however, the on-target editing is lower than for plasmid delivery (Figure 100). Protein delivery of BE3 with the HF mutations installed reduces on-targeting ediing efficiency but still yields some edited cells (Figure 101).

[00516] Both lipofection and installing HF mutations were shown to decrease off-target deamination events. For the four sites shown in Figure 102, the off-target sitest (OT) with the highest GUIDE-Seq reads and deamination events were assayed (Komor *et al., Nature,* 2016). The specificity ratio was calculated by dividing the off-target editing by the on-target editing at the closest corresponding C. In cases where off-target editing was not detectable, the ratio was set to 100. Thus, a higher specificity ratio indicates a more specific construct. BE3 plasmid delivery showed much higher off-target/on-target editing than protein delivery of BE3, plasmid delivery of HF-BE3, or protein delivery of HF-BE3 (Figures 102 and 105).

[00517] Purified proteins HF-BE3 and BE3 were analyzed *in vitro* for their capabilities to convert C to T residues at different positions in the spacer with the most permissive motif. Both

BE3 and HF-BE3 proteins were found to have the same "window" for base editing (Figures 103 and 104).

[00518] A list of the disease targets is given in Table 8. The base to be edited in Table 8 is indicated in bold and underlined.

Table 8. Base Editor Disease Targets

GENE	DISEASE	SPACER	SEQ ID NO	PAM	EDITOR	DEFECT	CELL
RB1	RETINOBLASTOMA	AAT**C**TAGTAAATAAATTGATGT	571	AAAAGT	SAKKH-BE3	SPLICING IMPAIRMENT	J82
PTEN	CANCER	GACCAA**C**GGCTAAGTGAAGA	572	TGA	VQR-BE3	W111R	MC116
PIK3CA	CANCER	TC**C**TTTCTTCACGGTTGCCT	573	ACTGGT	SAKKH-BE3	K111R	CRL-5853
PIK3CA	CANCER	CTC**C**TGCTCAGTGATTTCAG	574	AGA	VQR-BE3	Q546R	CRL-2505
TP53	CANCER	TGT**C**ACACATGTAGTTGTAG	575	TGG	YEE-BE3	N239D	SNU475
HRAS	CANCER	CCTCC**C**GGCCGGCGGTATCC	576	AGG	YEE-BE3	Q61R	MC/CAR

Table 6. Exemplary diseases that may be treated using base editors. The protospacer and PAM sequences (SEQ ID NOS: 577-589) are shown in the sgRNA (PAM) column. The PAM sequence is shown in parentheses and with the base to be edited indicated by underlining.

Disease target	gene symbol	Base changed	sgRNA (PAM)	Base editor
Prion disease	PRNP	R37*	GGCAGCCGATACCCGGGGCA(GGG)	BE3
			GGGCAGCCGATACCCGGGGC(AGG)	
Pendred syndrome	Slc26a4	c.919-2A>G	TTATTGTCCGAAATAAAAGA(AGA)	BE3
			ATTGTCCGAAATAAAAGAAG(AGG)	(VQR
			TTGTCCGAAATAAAAGAAGA(GGA)	SaCas9)
			GTCCGAAATAAAAGAAGAGGAAAA(AAT)	
			GTCCGAAATAAAAGAAGAGGAAAAA(ATT)	
Congenital deafness	Tmc1	c.545A>G	CAGGAAGCACGAGGCCACTG(AGG)	BE3
			AACAGGAAGCACGAGGCCAC(TGA)	YE-BE3
			AGGAAGCACGAGGCCACTGA(GGA)	YEE-BE3
Acquired deafness	SNHL	S33F	TTGGATTCTGGAATCCATTC(TGG)	BE3
Alzheimer's Disease	APP	A673T	TCTGCATCCATCTTCACTTC(AGA)	BE3 VQR
Niemann-Pick Disease Type C	NPC1	I1061T	CTTACAGCCAGTAATGTCAC(CGA)	BE3 VQR

EXAMPLE 14: Testing Base Editing Constructs

[00519] Several base editing constructs, including BE3, BE4-pmCDAl, BE4-hAID, BE4-3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE4-XTEN, BE4-32aa, BE4-2xUGI, and BE4 were tested for their ability to edit a cytosine (C) residue within different target sequences *{i.e.*, EMX1, FANCF, HEK2, HEK3, HEK4, and RNF2). For example, it was tested whether these constructs were capable of producing a C to T mutation. Schematic representations of the base editing constructs are shown in Figure 109. The target sequences tested are also shown in Figure 109 with the targeted cytosine numbered and indicated in red.

[00520] The following amino acid sequences were used in the base editing constructs of this example:

UGI:

TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAP EYKPWALVIQDSNGENKIKML (SEQ ID NO: 736)

rAPOBECl:

SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKH VEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHA DPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLEL YCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 737)

pmCDAl:

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQ SGTERGIH AEIFS IRKVEEYLRDNPGQFTINW YSSWSPC ADC AEKILEWYNQELRGNGHT LKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENR WLEKTLKRAEKRRS ELS IMIQ VKILHTTKS PAV (SEQ ID NO: 81)

hAID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHV ELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFC EDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSR QLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 49)

hAPOBEC3G:

MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTL TIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEP WNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTW VLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSC AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTF VDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 60)

SSB (single-stranded DNA binding protein):

ASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKATGEMKEQTEWHRV VLFGKLAEVASEYLRKGSQVYIEGQLRTRKWTDQSGQDRYTTEVVVNVGGTMQMLGG RQGGGAPAGGNIGGGQPQGGWGQPQQPQGGNQFSGGAQSRPQQSAPAAPSNEPPMDF DDDIPF (SEQ ID NO: 590)

Linker Sequences:

XTEN:

SGSETPGTSESATPES (SEQ ID NO: 604)

32aa:

SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 605)

SGGS:

SGGS (SEQ ID NO: 606)

$(GGS)_3$:

GGSGGSGGS (SEQ ID NO: 610)

[00521] The amino acid sequences of the constructs shown in Figure 109 are set forth below:

[00522] BE3:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESAT PESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN

SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDF YPFLKDNREKIEKILTFRIP YYVGPLARGNS RFAWMTRKS EETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDS PTVAY SVLVVAKVEKGKS KKLKS VKELLGITIMERS SFEKNPIDFLE AKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS
GGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS
DAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 174)

BE4-pmCDAl:

[00523] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGY
AVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQEL
RGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQ
LNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPESDKKYSIG
LAIGTNS VGW AVITDE YKVPS KKFKVLGNTDRHS IKKNLIG ALLFD SGET AEATRLKRT A
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL

VQTYNQLFEENPIN ASGVDAKAILS ARLS KSRRLENLIAQLPGEKKNGLFGNLIALS LGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS IPHQIHLGELHAILRRQED
FYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA
QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRK
LINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR
KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV
AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLP
KYSLFELENGRKRML ASAGELQKGNELALPS KYVNFLYLAS HYEKLKGS PEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIE
KETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA
LVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 175)

[00524] BE4-hAID:
MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHV
ELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFC
EDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSR
QLRRILLPLYE VDDLRD AFRTLGLS GSETPGTS ESATPES DKKYS IGLAIGTNS VGW AVIT
DEYKVPS KKFKVLGNTDRHS IKKNLIGALLFDS GETAE ATRLKRTARRRYTRRKNRIC Y
LQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI

NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGG ASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIE
KILTFRIP YYVGPLARGNS RFAWMTRKS EETITPWNFEE VVDKG ASAQS FIERMTNFDKN
LPNEKVLPKHS LLYE YFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKS KKLKS VKELLGITIMERS SFEKNPIDFLE AKGYKE VKKDLIIKLPKYS LFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESIL
MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIK
MLS GGS PKKKRKV (SEQ ID NO: 176)

[00525] BE4-3G:
MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTL
TIFVARLY YFWDPD YQE ALRS LCQKRDGPRATMKIMNYDEFQHCWS KFVYS QRELFEP
WNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTW
VLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSC
AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTF
VDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQENSGSETPGTSESATPESDKKYSIGLAI
GTNS VGW A VITDEYKVPS KKFKVLGNTDRHS IKKNLIGALLFDS GETAE ATRLKRTARR

RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ
TYNQLFEENPIN ASGVDAKAILS ARLS KSRRLENLIAQLPGEKKNGLFGNLIALS LGLTPN
FKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ
EEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY
PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV
DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKET
GKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVI
QDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 177)

[00526] BE4-N:
MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
APEYKPWALVIQDSNGENKIKMLGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKE
TCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGE
CSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNF
VNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQ

RLPPHILW ATGLKS GSETPGTS E SATPES DKKYS IGLAIGTNS VGW A VITDE Y KVPS KKF
KVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK
VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR
LIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS
ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD
DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGAS QEEFYKFIKPILEKMDGTEELLVKL
NREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNS RFA WMTRKS EETITPWNFEE VVDKGA SAQS FIERMTNFD KNLPNEKVLPKHS LL
YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI
ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDS LTFKEDIQKAQVS GQGDS LHEHIANLAGSPAIKKGILQT VKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL
YLY YLQNGRDM YVDQELDINRLS DYDVDHI VPQS FLKDDS IDNKVLTRS DKNRGKS DN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLES EFVYGDYKV YDVRKMIAKS EQEIGKAT AKYFFYS NIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQS ITGLYETRIDLS QLGGDS GGSPKKKRKV (SEQ ID NO: 178)

[00527] BE4-SSB:
MSSETGPVA VDPTLRRRIEPHEFE VFFDPRELRKETCLLYEINWGGRHS IWRHTS QNTNK
HVEVNFIEKFTTERYFCPNTRC SrrWFLSW SPCGEC SRAITEFLS RYPH VTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESAT
PESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET

AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDF YPFLKDNREKIEKILTFRIP YYVGPLARGNS RFAWMTRKS EETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDS PTVAY SVLV VAKVEKGKS KKLKS VKELLGITIMERS SFEKNPIDFLE AKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS
GGSGGSGGSASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKATGEM
KEQTEWHRVVLFGKLAEVASEYLRKGSQVYIEGQLRTRKWTDQSGQDRYTTEVVVNV
GGTMQMLGGRQGGGAPAGGNIGGGQPQGGWGQPQQPQGGNQFSGGAQSRPQQSAPA
APS NEPPMDFDDDIPFS GGS PKKKRKV (SEQ ID NO: 179)

[00528] BE4-$(GGS)_3$:
MSSETGPVA VDPTLRRRIEPHEFE VFFDPRELRKETCLLYEINWGGRHS IWRHTS QNTNK
HVEVNFIEKFTTERYFCPNTRC SrrWFLSWSPCGECSRAITEFLS RYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL

ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESAT
PESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDF YPFLKDNREKIEKILTFRIP YYVGPLARGNS RFAWMTRKS EETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDS PTVAY SVLV VAKVEKGKS KKLKS VKELLGITIMERS SFEKNPIDFLE AKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS
GGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDEN
VMLLTS DAPE YKPW ALVIQDS NGENKIKMLS GGS PKKKRKV (SEQ ID NO: 180)

[00529] BE4-XTEN:

MS SETGPVA VDPTLRRRIEPHEFE VFFDPRELRKETCLLYEINWGGRHS IWRHTS QNTNK
HVEVNFIEKFTTERYFCPNTRC SrrWFLS W SPCGEC SRAITEFLS RYPH VTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL

ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESAT
PESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDF YPFLKDNREKIEKILTFRIP YYVGPLARGNS RFAWMTRKS EETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDS PTVAY SVLV VAKVEKGKS KKLKS VKELLGITIMERS SFEKNPIDFLE AKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS
GSETPGTSESATPESTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDE
STDEN VMLLTS DAPE YKPW ALVIQDS NGENKIKMLS GGS PKKKRKV (SEQ ID NO: 181)

[00530] BE4-32aa:

MS SETGPVA VDPTLRRRIEPHEFE VFFDPRELRKETCLL YEINWGGRHS IWRHTS QNTNK
HVEVNFIEKFTTERYFCPNTRC SrrWFLS W SPCGEC SRAITEFLS RYPH VTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL

ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKS GGSSGGSSGSET
PGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ
LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA
NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK
YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILV
HTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID
NO: 182)

[00531] BE4-2XUGI:
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSrrWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH

ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESAT
PESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDF YPFLKDNREKIEKILTFRIP YYVGPLARGNS RFAWMTRKS EETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDS PTVAY SVLV VAKVEKGKS KKLKS VKELLGITIMERS SFEKNPIDFLE AKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS
GGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS
DAPEYKPWALVIQDSNGENKIKMLSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKR
KV (SEQ ID NO: 183)

[00532] BE4:

MSSETGPVAVDPTLRRRIEPHEFE VFFDPRELRKETCLLYEINWGGRHS IWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRC SrrWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKS GGSSGGSSGSET
PGTSESATPESSGGSSGGSDKKYSIGLAIGTNS VGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ
LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA
NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK
YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKP
ESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTN
LSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEY
KPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 184)

[00533] The ability of the base editing constructs of Figure 109 to mutate target cytosine residues of EMX1, FANCF, HEK2, HEK3, HEK4, and RNF2 are shown in Figures 110-115. The percentage of target cytosines edited (including the proportion of C residues that are mutated to a T, A, or G), as well as the % of indels generated, are shown in Figures 110-115. The percentage of target cytosines edited were calculated using the following formula: 100-[% of sequencing reads with C]. The pie charts of Figures 110-1 15 show the distribution of reads with the various bases indicated, meaning that looking at all of the base edited reads (reads that have a nucleotide other than a C at the base indicated), what percentage of those have A, G, or T. Tables 9-14, below, show the values of the mutation percentages that are indicated by in the pie charts of Figures 110-115.

[00534] The C to non-T editing observed is likely due to UDG (uracil DNA glycosylase). For example, once the C is converted to the uracil intermediate, UDG can convert it to an abasic site. This abasic site is then processed by other endogenous enzymes and ultimately leads to indels or other bases (such as G or A) replacing the C. We have shown that in UDG knock-out cell lines show increased C to T editing with little to no indels at all.

TABLE 9 EMX1

EMX1		C_5
BE3	A	1.9%
	C	51.9%
	G	6.0%
	T	40.3%
pmCDAl	A	0.2%
	C	88.9%
	G	0.8%
	T	10.0%
hAI D	A	0.3%
	C	83.6%
	G	0.6%
	T	15.4%
hAPOBEC3G	A	0.0%
	C	98.9%

	G	0.2%
	T	1.0%
BE4-N	A	1.4%
	C	85.6%
	G	3.5%
	T	9.5%
EMX1		
BE4-SSB	A	0.3%
	C	96.7%
	G	0.7%
	T	2.3%
BE4-$(GGS)_3$	A	1.7%
	C	36.8%
	G	4.1%
	T	57.4%
BE4-XTEN	A	2.1%
	C	45.7%
	G	5.1%
	T	47.1%
BE4-32aa	A	1.6%
	C	46.3%
	G	4.6%
	T	47.5%
BE4-2XUGI	A	0.6%
	C	60.2%
	G	1.9%
	T	37.3%

TABLE 10 FANCF

FANCF		C_8
BE3	A	1.6%
	C	74.4%
	G	0.7%
	T	23.3%
pmCDAl	A	0.5%
	C	88.8%
	G	0.3%
	T	10.3%

hAI D	A	0.3%
	C	89.6%
	G	0.2%
	T	9.9%
hAPOBEC3G	A	2.1%
	C	75.4%
	G	10.4%
	T	12.1%
BE4-N	A	1.8%
	C	83.5%
	G	0.6%
	T	14.2%
BE4-SSB	A	0.3%
	C	97.1%
	G	0.1%
	T	2.6%
BE4-$(GGS)_3$	A	2.3%
	C	45.7%
	G	1.4%
	T	50.6%
BE4-XTEN	A	1.1%
	C	56.3%
	G	0.4%
	T	42.1%
BE4-32aa	A	1.2%
	C	70.0%
	G	0.6%
	T	28.2%
BE4-2XUGI	A	0.9%
	C	57.8%
	G	1.1%
	T	40.2%

TABLE 11 HEK2

HEK2		C_6
BE3	A	0.9%

	C	28.2%
	G	52.9%
	T	18.1%
pmCDAl	A	1.8%
	C	73.5%
	G	3.5%
	T	21.2%
hAI D	A	2.1%
	C	56.9%
	G	7.7%
	T	33.3%
hAPOBEC3G	A	0.1%
	C	86.5%
	G	9.9%
	T	3.5%
BE4-N	A	1.0%
	C	37.6%
	G	57.0%
	T	4.4%
BE4-SSB	A	0.2%
	C	78.4%
	G	20.0%
	T	1.4%
BE4-$(GGS)_3$	A	0.6%
	C	11.1%
	G	40.6%
	T	47.7%
BE4-XTEN	A	1.2%
	C	24.8%
	G	44.6%
	T	29.4%
BE4-32aa	A	1.1%
	C	26.3%
	G	41.8%
	T	30.7%
BE4-2XUGI	A	0.8%

	C	37.0%
	G	21.6%
	T	40.6%

TABLE 12 HEK3

HEK3		C_5
BE3	A	2.23%
	C	38.06%
	G	12.77%
	T	46.95%
pmCDAl	A	0.21%
	C	76.57%
	G	0.12%
	T	23.09%
hAI D	A	0.28%
	C	60.23%
	G	1.03%
	T	38.45%
hAP0BEC3G	A	3.11%
	C	33.89%
	G	28.59%
	T	34.41%
BE4-N	A	2.6%
	C	64.1%
	G	13.5%
	T	19.8%
BE4-SSB	A	0.4%
	C	92.9%
	G	2.8%
	T	3.9%
BE4-$(GGS)_3$	A	1.3%
	C	9.9%
	G	7.9%
	T	80.8%
BE4-XTEN	A	2.3%
	C	15.9%
	G	12.2%

	T	69.6%
BE4-32aa	A	1.3%
	C	14.9%
	G	9.9%
	T	73.9%
BE4-2XUGI	A	0.6%
	C	23.4%
	G	3.8%
	T	72.2%

TABLE 13 HEK4

HEK4		c_5
BE3	A	8.40%
	C	41.89%
	G	24.54%
	T	25.17%
pmCDAl	A	0.50%
	C	87.53%
	G	0.01%
	T	11.95%
hAI D	A	0.93%
	C	71.32%
	G	0.69%
	T	27.06%
hAP0BEC3G	A	0.12%
	C	99.37%
	G	0.35%
	T	0.16%
BE4-N	A	7.3%
	C	56.6%
	G	25.7%
	T	10.3%
BE4-SSB	A	2.1%
	C	86.8%
	G	5.8%
	T	5.2%

BE4-$(GGS)_3$	A	6.7%
	C	13.0%
	G	19.8%
	T	60.5%
BE4-XTEN	A	7.5%
	C	19.7%
	G	25.4%
	T	47.4%
BE4-32aa	A	7.9%
	C	21.8%
	G	25.1%
	T	45.3%
BE4-2XUGI	A	3.4%
	C	22.2%
	G	12.4%
	T	62.0%

TABLE 14 RNF2

RNF2		C_6
BE3	A	2.46%
	C	46.65%
	G	19.87%
	T	31.03%
pmCDAl	A	0.60%
	C	83.52%
	G	1.33%
	T	14.55%
hAI D	A	0.36%
	C	75.03%
	G	3.20%
	T	21.40%
hAP0BEC3G	A	0.10%
	C	86.60%
	G	3.70%
	T	9.59%
BE4-N	A	5.1%
	C	50.0%

	G	**28.8%**
	T	**16.2%**
BE4-SSB	**A**	**1.1%**
	C	**89.9%**
	G	**4.9%**
	T	**4.1%**
BE4-$(GGS)_3$	**A**	**2.0%**
	C	**23.0%**
	G	**14.0%**
	T	**61.0%**
BE4-XTEN	**A**	**2.6%**
	C	**32.4%**
	G	**16.0%**
	T	**49.0%**
BE4-32aa	**A**	**2.2%**
	C	**29.2%**
	G	**18.5%**
	T	**50.0%**
BE4-2XUGI	**A**	**0.7%**
	C	**45.0%**
	G	**6.5%**
	T	**47.8%**

EXAMPLE 15: Base Editors Comprising a Cpfl Nickase that Cleaves the Targeted Strand.

[00535] As discussed above, nucleic acid programmable DNA binding proteins (napDNAbp) of any of the fusion proteins provided herein may be a Cpfl protein. In some embodiments, the Cpfl protein is a Cpfl nickase (nCpfl). Cpfl nickases, for example, a Cpfl nickase (R1225A in AsCpfl; and R1138A in LbCpfl) that cleaves the non-target strand have been described in Yamano *et al.,* "Crystal structure of Cpfl in complex with guide RNA and target DNA." Cell (165) 2016, p. 949-962; the entire contents of which is hereby incorporated by reference. However, a nickase *(e.g.,* a Cpfl nickase of a base editor protein) that cleaves the target strand is expected to improve base editing efficiency.

[00536] A fluorescent labeled DNA was used to identify a Cpfl mutant that preferentially nicks the target strand, rather than the non-target strand (see Figure 116). In Figure 116, the top

strand of DNA constructs 1-3, is the non-target strand and the bottom strand is the target strand. An *in vitro* assay is carried out using wild-type LbCpfl, R836A (LbCpfl), and R1138A (LbCpfl). R836A (LbCpfl) appears to be a *"crippled"* nickase, meaning it cuts the target strand more efficiently than the non-target strand. As shown in Figure 117, the non-target strand is uncut, no fluorescent 350 piece is observed. After two hours, both strands are cut. Differing intensities suggest more target strands are cut than non-target strands.

Establishing a Base Editing Window with AsCpl -BE3

[00537] Base editing proteins *(e.g.,* BE3 (SpCas9-BE3)) having LbCpfl(R836A) or AsCpfl(R912A) as the napDNAbp were shown to edit bases at low efficiency (0.1% to 0.4%). A base editor with a AsCpfl (R912A) napDNAbp more efficiently mutated a target C at EMX1, FANCF, HEK3 and HEK4 sites. The editing window of the constructs tested appears to be from the 7th base to the 11th base. The numbers are consistent with the trend with BE3 having highest numbers and self-defeating BE *(i.e.,* APOBEC-AsCpfl(R1225A)-UGI, which cleaves the non-target strand) having lower ones. See Figure 118 Positive control with Cas9-BE3 on EMX1: 5-6%. Indel values for AsCpfl: >20%. R912 in AsCpfl is conserved across many members of the Cpf 1 family. The corresponding residue in LbCpfl is R836, which is believed to be a *"crippled"* nickase when the R is mutated to an A.

Optimization of Cpf1-BE (Linkers)

[00538] Indel data suggests that Cpfl can access DNA target sites. Thus, optimization of Cpfl base editing proteins has focusd on specific APOBEC proteins, linkers, and/or UGI domains. The construct shown in Figure 119 was tested, with varying linkers using both LbCpfl (R836A) and AsCpfl (R912A). In short, different linker sequences *(i.e.,* XTEN, GGS, $(GGS)_3$ (SEQ ID NO: 610), $(GGS)_5$ (SEQ ID NO: 610), and $(GGS)_7$ (SEQ ID NO: 610)) between the APOBEC and Cpfl domain *(e.g.,* AsCpfl or LbCpfl) were tested. See Figure 120. The constructs were tested for their ability to mutate the C_8 residue of the HEK3 site, which is TGCTTCTCsCAGCCCTGGCCTGG (SEQ ID NO: 592). Editing levels for base editing proteins with AsCpfl reached to over 1%, while base editing proteins with LbCpfl showed a comparative reduction in base editing efficiency. As shown in Figure 121, linkers from a database maintained by the Centre of Integrative Bioinformatics VU did not show as significant

an improvement as GGS-type linkers for AsCpfl-BE3. The linkers shown in Figure 121 are shown below:

[00539]	PDB_code	Length (aa)	Sequence
[00540]	lau7A_1	10	KRRTTISIAA (SEQ ID NO: 593)
[00541]	lclkA_1	19	ALVFYREYIGPvLKQIKFKF (SEQ ID NO: 594)
[00542]	lc20A_1	14	LPIMAKSVLDLYEL (SEQ ID NO: 595)
[00543]	lee8A_1	5	LLRLG (SEQ ID NO: 596)
[00544]	lflzA_1	15	TDKEINPVVKENIEW (SEQ ID NO: 597)
[00545]	lignA_1	8	PPSIKRKF (SEQ ID NO: 598)
[00546]	ljmcA_1	9	LPTVQFDFT (SEQ ID NO: 599)
[00547]	lsfe_l	14	LPLDIRGTAFQQQV (SEQ ID NO: 600)
[00548]	2ezx_l	8	AYVVLGQF (SEQ ID NO: 601)
[00549]	2reb_l	8	INFYGELV (SEQ ID NO: 602)

Optimization of Cpfl -BE (Orientations)

[00550] Cas9 has a stretch of amino acids between the C and N termini (see red square, Figure 123) while AsCpfl does not (see Figure 122). Moreover, AsCpfl has a shorter distance between the N and C termini. These observations indicate potential interference between APOBEC (on N terminus) and UGI (on C terminus) through which UGI may hinder APOBEC access to the non-target strand. One solution is to move APOBEC and UGI onto the same terminus, either N or C. Accordingly, constructs having the architecture NLS-UGI-APOBEC-XTEN-AsCpfl; UGI-APOBEC-XTEN-AsCpfl-NLS; and AsCpfl -XTEN-APOBEC-NLS will be tested.

Optimization of Cpfl -BE (Internal Truncation)

[00551] There is no known crystal structure of Cpfl in which the non-target strand is resolved (see Figure 124, cyan). It is believed that the editing window should lie within the red circle as shown in Figure 124. There is a helical region (see square in Figure 124) that may be obstructing APOBEC. This region comprises the amino acid sequence K(661)KTGDQK(667) (SEQ ID NO: 603).

[00552] To test the whether the removal of two, four or six residues improves base editing efficiency, experiments were conducted with a base editor having a AsCpfl(R912A) napDNAbp, using HEK3 as the target site. Editing levels increase to approximately 2.6% - a 6-fold increase from controll levels when T663 and D665 are deleted (see Table 7, below). The construct used in this experiment was APOBEC-XTEN-AsCpfl(R912A)-SGGS-UGI

Table 7

Deletions	Editing at C8	Editing at C9
T663, D665	2.59%	1.29%
K662, T663, D665, Q666	0.15%	0.15%
K661, K662, T663, D665, Q666, K667	0.22%	0.21%

REFERENCES

1. Humbert O, Davis L, Maizels N. Targeted gene therapies: tools, applications, optimization. *Crit Rev Biochem Mol.* **2012;** 47(3):264-81. PMID: 22530743.
2. Perez-Pinera P, Ousterout DG, Gersbach CA. Advances in targeted genome editing. *Curr Opin Chem Biol.* **2012;** 16(3-4):268-77. PMID: 22819644.
3. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet.* **2010;** ll(9):636-46. PMID: 20717154.
4. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol.* **2013;** 14(l):49-55. PMID: 23169466.
5. Charpentier E, Doudna JA. Biotechnology: Rewriting a genome. *Nature.* **2013;** 495, (7439):50-l. PMID: 23467164.
6. Pan Y, Xia L, Li AS, Zhang X, Sirois P, Zhang J, Li K. Biological and biomedical applications of engineered nucleases. *Mol Biotechnol.* **2013;** 55(l):54-62. PMID: 23089945.
7. De Souza, N. Primer: genome editing with engineered nucleases. *Nat Methods.* **2012;** 9(1):27. PMID: 22312638.
8. Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, CoUingwood TN. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci USA.* **2008;** 105(15):5809-14. PMID: 18359850.
9. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley

GQ, Lander ES. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet.* **1999;** 22(3):231-8. PMID: 10391209.

10. Jansen R, van Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol.* **2002;** 43(6): 1565-75. PMID: 11952905.

11. Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nat Methods.* **2013;** 10(10):957-63. PMID: 24076990.

12. Jore MM, Lundgren M, van Duijin E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Shou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol.* **2011;** 18(5):529-36. PMID: 21460843.

13. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science.* **2010;** 327(5962): 167-70. PMID: 20056882.

14. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature.* **2012;** 482(7385):331-8. PMID: 22337052.

15. Gasiunas G, Siksnys V. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol.* **2013;** 21(1 1):562-7. PMID: 24095303.

16. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* **2013;** 152(5): 1173-83. PMID: 23452860.

17. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods.* **2013;** 10(10):973-6. PMID: 23892895.

18. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* **2013;** 31(9):833-8. PMID: 23907171.

19. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* **2013;** 154(2):442-51. PMID: 23849981.

20. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc.* **2013;** 8(1 1):2180-96. PMID: 24136345.

21. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science.* **2013;** 339(6121):823-6. PMID: 23287722.

22. Cole-Strauss A, Yoon K, Xiang Y, Byrne BC, Rice MC, Gryn J, Holloman WK, Kmiec EB. Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide. *Science.* **1996;** 273(5280): 1386-9. PMID: 8703073.
23. Tagalakis AD, Owen JS, Simons JP. Lack of RNA-DNA oligonucleotide (chimeraplast) mutagenic activity in mouse embryos. *Mol Reprod Dev.* **2005;** 71(2):140-4. PMID: 15791601.
24. Ray A, Langer M. Homologous recombination: ends as the means. *Trends Plant Sci.* **2002;** 7(10):435-40. PMID 12399177.
25. Britt AB, May GD. Re-engineering plant gene targeting. *Trends Plant Sci.* **2003;** 8(2):90-5. PMID: 12597876.
26. Vagner V, Ehrlich SD. Efficiency of homologous DNA recombination varies along the Bacillus subtilis chromosome. *J Bacteriol.* **1988;** 170(9):3978-82. PMID: 3137211.
27. Saleh-Gohari N, Helleday T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res.* **2004;** 32(12):3683-8. PMID: 15252152.
28. Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol.* **2007;** 25(11): 1298-306. PMID: 17965707.
29. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **2008;** 9(6):229. PMID: 18598372.
30. Reynaud CA, Aoufouchi S, Faili A, Weill JC. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nat Immunol.* **2003;** 4(7):631-8.
31. Bhagwat AS. DNA-cytosine deaminases: from antibody maturation to antiviral defense. *DNA Repair (Amst).* **2004;** 3(l):85-9. PMID: 14697763.
32. Navaratnam N, Sarwar R. An overview of cytidine deaminases. *Int J Hematol.* **2006;** 83(3): 195-200. PMID: 16720547.
33. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, Stevens RC, Goodman MF, Chen XS. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature.* **2008;** 456(7218): 121-4. PMID: 18849968.
34. Chelico L, Pham P, Petruska J, Goodman MF. Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G. *J Biol Chem.* **2009;** 284(41). 27761-5. PMID: 19684020.
35. Pham P, Bransteitter R, Goodman MF. Reward versus risk: DNA cytidine deaminases triggering immunity and disease. *Biochemistry.* **2005;** 44(8):2703-15. PMID 15723516.
36. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev.* **2013;** 65(10): 1357-69. PMID: 23026637.
37. Lee JW, Soung YH, Kim SY, Lee HW, Park WS, Nam SW, Kim SH, Lee JY, Yoo NJ, Lee SH. PIK3CA gene is frequently mutated in breast carcinomas and hepatocellular carcinomas. *Oncogene.* **2005;** 24(8): 1477-80. PMID: 15608678.

38. Ikediobi ON, Davies H, Bignell G, Edkins S, Stevens C, O'Meara S, Santarius T, Avis T, Barthorpe S, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Hunter C, Jenkinson A, Jones D, Kosmidou V, Lugg R, Menzies A, Mironenko T, Parker A, Perry J, Raine K, Richardson D, Shepherd R, Small A, Smith R, Solomon H, Stephens P, Teaque J, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Reinhold W, Weinstein JN, Stratton MR, Futreal PA, Wooster R. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther.* **2006;** 5(11):2606-12. PMJX>: 17088437.

39. Cox, D. B., Piatt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nature medicine 21,* 121-131, doi:10.1038/nm.3793 (2015).

40. Hilton, I. B. & Gersbach, C. A. Enabling functional genomics with genome engineering. *Genome researchlZ,* 1442-1455, doi: 10. 1101/gr. 190124. 115 (2015).

41. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32,** 347-355, doi: 10.1038/nbt. 2842 (2014).

42. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology* **33,** 538-542, doi:10.1038/nbt.3190 (2015).

43. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology* **33,** 543-548, doi:10.1038/nbt.3198 (2015).

44. Lin, S., Staahl, B. T., Alia, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3,** e04766, doi:10.7554/eLife.04766 (2014).

45. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.l231143 (2013).

46. Rong, Z., Zhu, S., Xu, Y. & Fu, X. Homologous recombination in human embryonic stem cells using CRISPR/Cas9 nickase and a long DNA donor template. *Protein & cell* **5,** 258-260, doi: 10. 1007/s 13238-014-0032-5 (2014).

47. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuc lease in Adaptive Bacterial Immunity. *Science* **337,** 816-821, doi: 10. 1126/science. 1225829 (2012).

48. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA Editing Enzyme APOBECl and Some of Its Homologs Can Act as DNA Mutators. *Molecular Cell* **10,** 1247-1253 (2002).

49. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343,** 1247997, doi: 10.1 126/science. 1247997 (2014).

50. Schellenberger, V. *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nature biotechnology* **27,** 1186-1 190, doi:10.1038/nbt.l588 (2009).

51. Saraconi, G., Severi, F., Sala, C , Mattiuz, G. & Conticello, S. G. The RNA editing enzyme APOBECl induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome biology* **15,** 417- (2014).
52. Anders, C , Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569-573, doi:10.1038/naturel3579 (2014).
53. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
54. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR- Cas nucleases. *Nature biotechnology* **33,** 187-197, doi:10.1038/nbt.3117 (2015).
55. Kunz, C , Saito, Y. & Schar, P. DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cellular and molecular life sciences : CMLS* **66,** 1021-1038, doi: 10. 1007/sOOO 18-009- 8739-9 (2009).
56. D., M. C. *et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82,** 701-708 (1995).
57. Caldecott, K. W. Single-strand break repair and genetic disease. *Nature reviews. Genetics 9*, 619-631, doi:10.1038/nrg2380(2008).
58. Fukui, K. DNA mismatch repair in eukaryotes and bacteria. *Journal of nucleic acids* **2010,** doi:10.4061/2010/260512 (2010).
59. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109,** E2579-E2586, doi: 10. 1073/pnas. 1208507 109 (2012).
60. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520,** 186-191, doi:10.1038/naturel4299 (2015).
61. Kuscu, C , Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature biotechnology* **32,** 677-683, doi: 10. 1038/nbt.2916 (2014).
62. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology* **32,** 670-676, doi: 10. 1038/nbt.2889 (2014).
63. Beale, R. C. L. *et al.* Comparison of the Differential Context-dependence of DNA Deamination by APOBEC Enzymes: Correlation with Mutation Spectra in Vivo. *Journal of Molecular Biology* **337,** 585-596, doi:10.1016/j.jmb.2004.01.046 (2004).
64. Kim, J., Basak, J. M. & Holtzman, D. M. The role of apolipoprotein E in Alzheimer's disease. *Neuron* **63,** 287-303, doi: 10. 1016/j.neuron.2009.06.026 (2009).
65. Liu, C. C , Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature reviews. Neurology* **9,** 106- 118, doi:10.1038/nrneurol.2012.263 (2013).
66. Sjoblom, T. *et al.* The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *Science* **314,** 268-274, doi: 10. 1126/science. 1133427 (2006).

67. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486,** 400-404, doi: 10. 1038/naturel 1017 (2012).
68. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research,* doi:10.1093/nar/gkvl222 (2015).
69. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science,* doi:10.1126/science.aad5227 (2015).
70. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nature chemical biology* **11**, 316-318, doi:10.1038/nchembio.l793 (2015).
71. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature biotechnology* **33,** 73-80, doi:10.1038/nbt.3081 (2015).
72. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523,** 481-485, doi: 10. 1038/naturel 4592 (2015).
73. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA- programmed Cas9 nuclease specificity. *Nature Biotechnology* **31,** 839-843, doi:10.1038/nbt.2673 (2013).
74. Shcherbakova, D. M. & Verkhusha, V. V. Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nature Methods* **10**, 751-754, doi:10.1038/nmeth.2521 (2013).
75. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8,** 2281-2308, doi:10.1038/nprot.2013.143 (2013).
76. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science,* doi:10.1126/science.aad8282 (2016).
77. Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided Fokl nucleases for highly specific genome editing. *Nat Biotech* **32,** 569-576, doi:10.1038/nbt.2908 (2014).
78. Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* **4,** 712-720 (2003).
79. Heller, R. C. & Marians, K. J. Replisome assembly and the direct restart of stalled replication forks. *Nat Rev Mol Cell Biol* **7,** 932-943 (2006).
80. Pluciennik, A. *et al.* PCNA function in the activation and strand direction of MutLa endonuclease in mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 16066-16071, doi: 10. 1073/pnas. 1010662107 (2010).
81. Seripa, D. *et al.* The missing ApoE allele. *Annals of human genetics* **71,** 496-500, doi: 10.1 111/j. 1469- 1809.2006.00344.x (2007).
82. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529,** 490-495, doi:10.1038/naturel6526 (2016).
83. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotech* **34,** 339-344, doi:10.1038/nbt.3481 (2016).

84. Simonelli, V., Narciso, L., Dogliotti, E. & Fortini, P. Base excision repair intermediates are mutagenic in mammalian cells. *Nucleic acids research* **33,** 4404-4411, doi:10.1093/nar/gki749 (2005).
85. Barnes, D. E. & Lindahl, T. Repair and Genetic Consequences of Endogenous DNA Base Damage in Mammalian Cells. *Annual Review of Genetics* **38,** 445-476, doi:doi:10.1146/annurev.genet.38.072902.092448 (2004).

EXAMPLE 16: Improving DNA specificity and applicability of base editing through protein engineering and protein delivery

[00553] Base editing, a genome editing approach that enables the programmable conversion of one base pair into another without double-stranded DNA cleavage, excess stochastic insertions and deletions, or dependence on homology-directed repair was developed. The application of base editing is limited by off-target activity and reliance on intracellular DNA delivery. Here two advances are described that address these limitations. First, off-target base editing has been reduced by installing mutations into the third-generation base editor (BE3) to generate a high-fidelity base editor (HF-BE3). Next, BE3 and HF-BE3 are purified and delivered as ribonucleoprotein (RNP) complexes into mammalian cells, establishing DNA-free base editing. RNP delivery of BE3 confers higher specificity even than plasmid transfection of HF-BE3, while maintaining comparable on-target editing levels. Finally, these advances are applied to deliver BE3 RNPs into both zebrafish embryos and the inner ear of live mice to achieve specific, DNA-free base editing *in vivo.*

Introduction

[00554] Traditional genome editing agents introduce double-stranded DNA breaks (DSBs) as the first step of genome editing[1-4]. Cells respond to DSBs primarily through non-homologous end joining (NHEJ), resulting in stochastic insertions or deletions (indels) at the cleavage site[1,5]. To generate more precise changes in genomic DNA, homology-directed repair (HDR) can be used to replace the genomic DNA surrounding the cleavage site with that of an exogenously supplied DNA donor template[6-8]. Unfortunately, HDR is typically accompanied by an excess of indels resulting from competing NHEJ and is limited primarily to mitotic cells. In addition, most genome editing methods rely on delivery of exogenous plasmid or viral DNA into mammalian cells followed by intracellular expression of the agent[9-12]. These delivery methods result in

continuous, uncontrolled Cas9 and sgRNA expression even after the on-target locus has been edited, increasing the opportunity for genome editing at off-target loci[1,13].

[00555] Base editing, a different approach to genome editing that enables the direct, programmable, targeted conversion of a C:G base pair to a T:A base pair, was recently described[3,14]. The third-generation base editor, BE3, contains in a single protein (*i*) a catalytically impaired Cas9 that opens a small single-stranded DNA bubble at a guide RNA-specified locus, (*ii*) a tethered single-strand-specific cytidine deaminase that converts C to U within a window of approximately five nucleotides in the single-stranded DNA bubble, *(iii)* a uracil glycosylase inhibitor (UGI) that inhibits base excision repair, thereby improving the efficiency and product selectivity of base editing, and (iv) nickase activity to manipulate cellular mismatch repair into replacing the G-containing DNA strand. The combination of these components enables efficient and permanent C to T (or G to A) conversion in mammalian cells with minimal indel formation. Since prefiously reported[14], other researchers have confirmed the ability of this strategy and related approaches to facilitate Cas9-directed C to T conversion in mammalian cells[15-17] and in plants[18].

[00556] Here, two advances that greatly improve the DNA specificity of base editing and that allow base editing *in vitro* and *in vivo* without supplying exogenous DNA, which has been associated with a risk of recombination with the host genome and cytotoxicity, are described[18,19]. First, a mutant form of BE3 incorporating mutations known to decrease the DNA affinity of Cas9[20] that reduces off-target editing events with only a modest decrease in on-target editing activity is engineered. Next, it is revealed that lipid-mediated delivery of base editor proteins complexed with guide RNA results in even larger specificity enhancements with no apparent reduction in on-target base editing compared to plasmid DNA delivery. Delivery of base editors as RNPs typically reduces off-target editing to below measurable levels, even for a notoriously promiscuous guide RNA that targets a highly repetitive genomic DNA sequence, in cultured human and mouse cells. These advances enable highly specific, DNA-free *in vivo* base editing in mice and zebrafish to be demonstrated.

Results

Engineering a high-fidelity base editor

[00557] Cas9 nucleases and their associated fusion constructs have been shown to bind and cleave DNA at off-target genomic loci[21-24]. Joung and coworkers developed HF-Cas9, a high-fidelity SpCas9 variant containing four point mutations (N497A, R661A, Q695A, Q926A) that were designed to eliminate non-specific interactions between Cas9 and the phosphate backbone of the DNA target strand (Fig. 125A)[20] consistent with the previous abrogation of non-specific DNA interactions in TALENs that greatly increased their DNA cleavage specificity[25]. Since base editors operate on the non-target strand within the single-stranded DNA bubble created by Cas9[14] it can be hypothesized that introducing these four point mutations from HF-Cas9 into BE3 to generate "HF-BE3" might reduce off-target base editing without altering its base conversion capabilities (Figs. 125A and 125C).

[00558] Plasmids encoding BE3 and HF-BE3 as His_6-tagged proteins were overexpressed in *E. coli* and purified first by nickel affinity chromatography and then by cation exchange chromatography (Figs. 126A-126B). Following extensive optimization of expression and purification conditions, BE3 and HF-BE3 protein can be routinely produced at a yield of ~2 mg per liter of culture media (Figs. 126A-126C).

[00559] The purified base editor proteins were used to compare base editing efficiency and the width of the editing window of HF-BE3 and BE3 biochemically. *In vitro* C to U conversion efficiencies were measured in a synthetic dsDNA 79-mer with a protospacer comprised of TC repeats. The target dsDNA (250 nM) was incubated with BE3:sgRNA or HF-BE3:sgRNA (2 μM) for 30 min at 37 °C. After incubation, the edited DNA was amplified using a uracil-tolerant polymerase and sequenced by high-throughput DNA sequencing (HTS). Comparable editing efficiencies and activity window widths were observed for HF-BE3 and BE3 *in vitro* (Fig. 125B). These findings indicate that introduction of the high-fidelity mutations into BE3 does not compromise inherent on-target base editing efficiency or change the width of the editing window of the resulting HF-BE3 protein *in vitro.*

HF-BE3 enhances editing specificity following DNA transfection

[00560] Next, base editing efficiencies, specificities, and editing window widths of BE3 and HF-BE3 were compared in mammalian cells following plasmid DNA transfection. Four well-studied endogenous genomic loci (HEK293 site 3, FANCF, EMX1 and VEGFA site 2) were chosen to interrogate on- and off-target base editing in mammalian cells[14,24]. VEGFA site 2 is

highly repetitive, containing 14 Cs out of 20 protospacer nucleotides, and is associated with notoriously high rate of known off-target genome editing[20,22,24,26]. This site was chosen to be included because it poses a formidable specificity challenge. In contrast with most nuclease-based genome editing applications, base editing relies on the precise location of the protospacer to place the target nucleotide within the editing window and usually little or no flexibility in the choice of guide RNA is available. Therefore, the development of base editors with enhanced specificities even for highly repetitive, promiscuous sgRNA targets is crucial[3,14].

[00561] The on-target locus and known off-target loci were amplified by PCR and analyzed by HTS following plasmid transfection[24] with each of the four base editor: sgRNA pairs. On-target editing in HEK293T cells for these four endogenous genomic loci was slightly reduced by introduction of the HF mutations; editing averaged 29+5% with BE3, and 21+3% (mean ± s.e.m. for n = 3 biological replicates) for HF-BE3 (Figs. 127A-127D, 128A).

[00562] For each of the three standard, non-repetitive target sites (HEK293 site 3, FANCF, and EMX1), the three most frequently modified off-target loci that contain a C within the editing window from the off-target loci previously reported to be modified from treatment with Cas9 and the same guide RNA were examined (Table 15)[24]. When cells were transfected with BE3 plasmid, C-^T conversion across the nine most frequently modified Cas9 off-target loci for HEK293 site 3, FANCF, and EMX1 averaged 1.1+0.3% (Figs. 127A-C; mean ± s.d. for n = 3 biological replicates). Installation of the HF mutations reduced the absolute level of mean off-target editing by 37-fold to 0.03+0.005%, with only one instance of measureable off-target C-^T conversion (Fig. 127A; EMX1 C_5 at off-target 1).

Site	Sequence	SEQ ID NO	GUIDE-Seq count
EMX1 on-target	GAGTCCGAGCAGAAGAAGAAGGG	480	4,521
EMX1 off-target 1	GAGTCTAAGCAGAAGAAGAAGAG	481	1,445
EMXl off-target 2	GAGGC$_5$C$_6$GAGCAGAAGAAAGACGG	482	700
EMXl off-target 3	GAGTC$_5$C$_6$TAGCAGGAGAAGAAGAG	483	390
HEK293 site 3 on-target	**GGCC$_4$C$_5$AGACTGAGCACGTGATGG**	484	2,074
HEK293 site 3 off-target 1	**CACC $_4$C$_5$AGACTGAGCACGTGCTGG**	485	327

HEK293 site 3 off-target 2	GACAC$_5$AGACTGGGCACGTGAGGG	486	306
HEK293 site 3 off-target 3	AGCTC$_5$AGACTGAGCAAGTGAGGG	487	136
VEGFA site 2 on-target	GAC$_3$C$_4$C$_5$C$_6$C$_7$TC$_9$CioACCCCGCCTCC GG	488	540
VEGFA site 2 off-target 1	CTAC$_4$C$_5$C$_6$C$_7$TC$_9$CioACCCCGCCTCCG G	489	1,925
VEGFA site 2 off-target 2	ATTC$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCCCGCCTCAG **G**	490	1,549
VEGFA site 2 off-target 3	ACAC$_4$C$_5$C$_6$C$_7$C$_8$C$_9$CioACCCCGCCTCA GG	491	1,178
VEGFA site 2 off-target 4	TGC$_3$C$_4$C$_5$C$_6$C$_7$C$_8$C$_9$CioACCCCACCTCT GG	492	1,107
FANCF on-target	GGAATC$_6$C$_7$C$_8$TTC„TGCAGCACCTGG	493	4,816
FANCF off-target 1	GGAAC$_5$C$_6$C$_7$C$_8$GTC„TGCAGCACCAG G	494	2,099
FANCF off-target 2	GGAGTC$_6$C$_7$C$_8$TCioC„TACAGCACCAG **G**	495	524
FANCF off-target 3	AGAGGC$_6$C$_7$C$_8$C$_9$TC„TGCAGCACCAG G	496	150

Table 15: Protospacer and PAM sequences for the on- and off-target human genomic loci studied in this work. The off-target sites were chosen based on their GUIDE-Seq read count [45]. Cytosines within the editing window for a particular sgRNA are numbered. The PAM sequence is shown in bold. Protospacer bases in off-target loci that differ from their respective on-target loci have been underlined. For genomic sequences interrogated in murine samples, see Figure 132E.

[00563] To characterize HF-BE3 specificity on an extremely challenging site, BE3 and HF-BE3 off-target activity when targeting the highly repetitive VEGFA site 2 locus was compared. BE3 treatment lead to an average of 15+5% editing of cytosines located in the activity windows of the four tested off-target sites associated with this sgRNA (all average values quoted in this paragraph represent mean ± s.d. for n = 3 biological replicates). In contrast, HF-BE3 lead to a 3-fold reduction in absolute off-target editing (5.0+2.3%) at the same off-target sites (Fig. 127D). When compared to transfection of BE3, HF-BE3 significantly ($p < 0.05$, two-tailed Student's t test) reduced off-target editing at 27 of the 57 cytosines located at off-target loci (Table 16), while HF-BE3 treatment lead to a significant reduction ($p < 0.05$ two-tailed Student's t test) in on-target editing at only 3 of 16 the interrogated on-target cytosine residues.

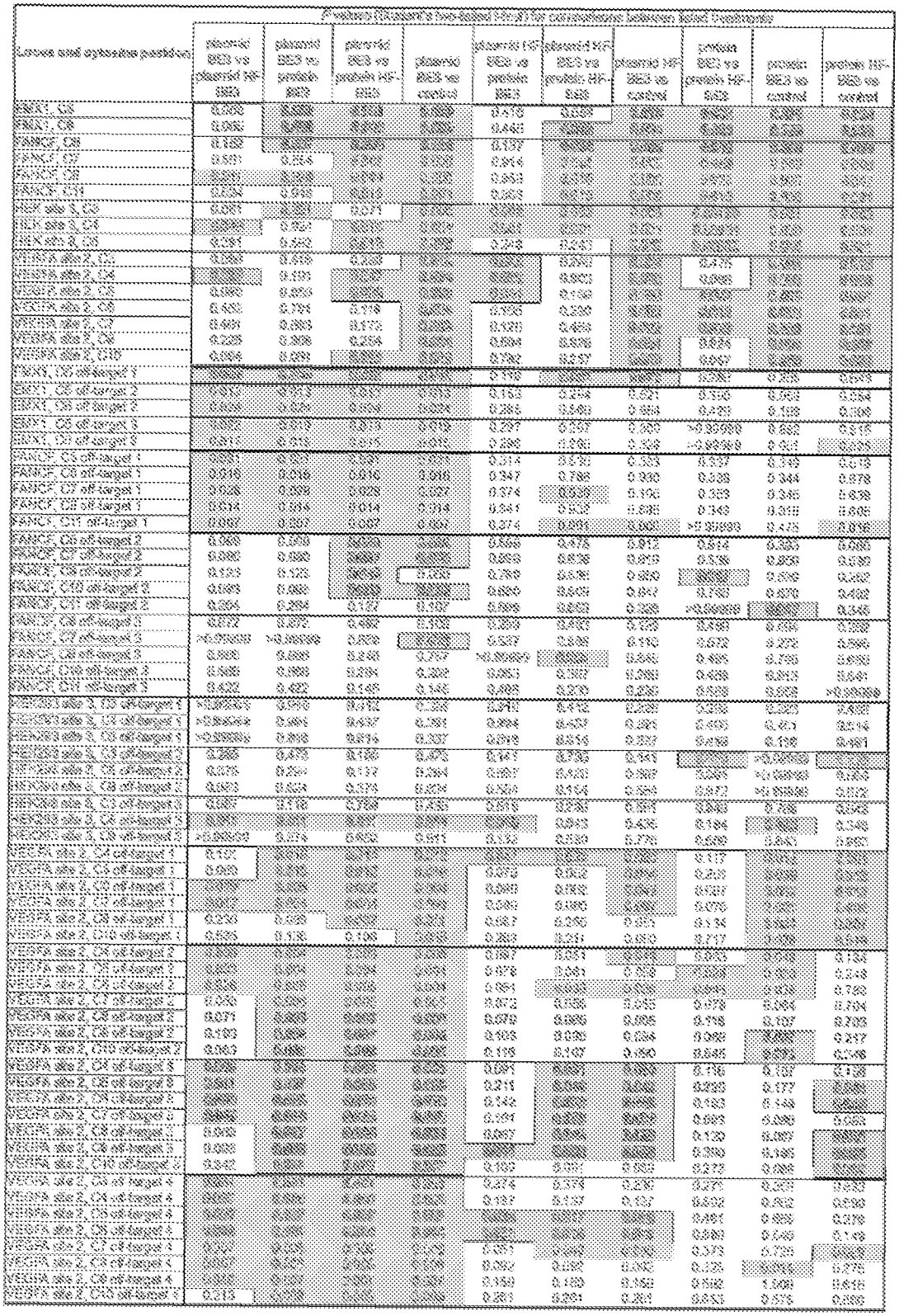


Table 16: *P*-values for differences in base editing under different treatment conditions at all loci evaluated in this study. p-values were calculated using the Student's two tailed t-test as described

in the Materials and Methods. When the p-value indicated a significant difference ($p < 0.05$), the corresponding entry has been highlighted.

[00564] In addition to considering the differences between absolute editing at off-target loci, the on-target:off-target editing specificity ratio was also calculated by dividing the observed on-target efficiency by the off-target efficiency (Figs. 129A-129B). This metric takes into account any reduction in on-target editing associated with installation of the HF-mutations, and is useful for applications sensitive to both the efficiency and specificity of base editing. Off-target editing by HF-BE3 was below the detection limit of high-throughput sequencing for several off-target loci. For these cases, a conservative off-target editing efficiency equal to the upper limit of detection was assumed (0.025% C-^T conversion; see Methods). Based on this analysis, the average improvement in specificity ratio upon installation of the HF mutations across all 34 target cytosines examined herein was 19-fold, when plasmid delivery of the two constructs was performed. These results collectively establish that for non-repetitive sites (Fig. 129A) as well as a highly repetitive site (Fig. 129B), HF-BE3 results in substantially enhanced base editing specificity with only a modest reduction in on-target editing efficiency compared to BE3.

RNP delivery ofBE3 enables DNA-free base editing

[00565] Next, the ability of BE3 in DNA-free, RNP form to mediate base editing when directly delivered into cultured human cells was studied. It has recently been established that cationic lipid reagents can potently deliver negatively charged proteins or protein: nucleic acid complexes into mammalian cells including ribonucleoprotein (RNP) complexes and that RNP delivery can substantially reduce off-target genome editing[27-29]

[00566] The commercially available cationic lipid Lipofectamine 2000 was combined with either purified BE3 protein or HF-BE3 protein after pre-complexation with a guide RNA targeting the EMX1, HEK293 site 3, FANCF, or VEGFA site 2 locus and the resulting lipid:RNP complexes were incubated with HEK293T cells. After 72 h, genomic DNA was harvested and on-target and off-target base editing was analyzed by high-throughput DNA sequencing. As with all Cas9-based technologies, substantial variations were observed in editing efficiency at different genomic loci (Figs. 127 and 130). To display trends associated with in on-

target editing efficiency between different treatments, mean on-target base editing efficiencies were calculated at the four tested loci (Fig. 128A). Protein delivery of BE3 (200 nM) lead to on-target editing efficiencies comparable to those observed with plasmid transfection (26+4% vs. 29+5% respectively; mean ± s.e.m. for n = 3 biological replicates; Fig. 128A).

[00567] In contrast, protein delivery of HF-BE3 reduced on-target editing compared to protein delivery of BE3 at the four genomic loci studied (average editing efficiency of 13+3% vs. 26+4%, respectively; mean ± s.e.m. for n = 3 biological replicates; Fig. 128A). Since HF-BE3 and BE3 have comparable editing efficiencies in a test tube (Fig. 125B) and editing is only slightly reduced when HF-BE3 is expressed from plasmids in HEK293T cells (Fig. 127A-D), it is tempting to speculate that the decreased efficiency of editing from HF-BE3 protein delivery may be a result of decreased HF-BE3 stability in mammalian cells. Lower stability could be offset by continual expression from a plasmid, but not following one-time protein delivery. This observation is consistent with a recent report of reduced on-target indel formation with purified HF-Cas9 compared to purified Cas9 when nucleofected into $CD34^+$ hematopoietic stem and progenitor cells[30]. While this work was in review, Kim et al demonstrated RNP delivery of BE3 into mouse embryos using electroporation[31]. To the best of the inventors' knowledge, the present approach is the first DNA-free technique capable of generating precise changes to individual nucleotides in mammalian cells without electroporation, which has limited *in vivo* therapeutic relevance.

RNP delivery of base editors greatly enhances DNA specificity

[00568] Importantly, while RNP delivery of BE3 and HF-BE3 led to substantial on-target base editing, no instances of measurable base editing (< 0.025%) were observed at any of the nine tested off-target loci associated with EMXl, FANCF and HEK293 site 3, (Figs. 130A-130C). In contrast, plasmid delivery of BE3 lead to an average of 1.1+0.3% (mean ± s.d. for n = 3 biological replicates) off-editing across all sequenced cytosines within the base editing activity window, and detectable off-target editing at 11 of 16 off-target cytosines located at these nine off-target loci (Figs. 127A-127D). At off-target loci of the three non-repetitive loci tested, BE3 protein delivery lead to a 26-fold higher average specificity ratio than that of plasmid delivery (Fig. 127A). These results reveal that RNP delivery of base editors dramatically increases the DNA specificity of base editing.

[00569] Protein delivery of either BE3 or HF-BE3 also resulted in greatly improved base editing specificity at the highly promiscuous VEGFA site 2 locus compared to plasmid delivery of either BE3 or HF-BE3 (compare Figs. 127 and 130; see Table 16). Absolute frequencies of base editing at the off-target loci associated with this site were reduced upon protein delivery at least 10-fold for both BE3 (plasmid delivery: 15+4% off-target editing; protein delivery: 1.3+0.4% off-target editing; all values in this paragraph represent mean ± s.d. for n = 3 biological replicates) and HF-BE3 (plasmid delivery: 5+2% off-target editing; protein delivery: 0.5+0.1% off-target editing). Across all four studied loci, base editing specificity ratios for on-target:off-target editing increased an average of 66-fold for protein delivery of BE3 compared with plasmid delivery of BE3 (Fig. 129). Collectively, these results reveal that for both repetitive and non-repetitive target sites, RNP versus DNA delivery is a stronger determinant of base editing specificity than the presence or absence of the high-fidelity Cas9 mutations.

[00570] Neither introduction of the HF mutations nor delivery method substantially altered the low indel rates associated with base editing. Indel frequencies at all on-target loci across all treatment conditions in this study remained low (typically ≤ 5%; Fig. 131A), and the editingdndel ratio remained higher in all cases tested (typically ≥ 10-fold; Fig. 13IB) than in previous studies using optimized HDR protocols[30,32,33]. For non-repetitive sgRNAs, very few indels were observed at off-target loci (Fig. 131C), although it is noted that plasmid delivery of BE3 generated up to 5% indels for off-target loci associated with VEGFA site 2 (Fig. 131C).

[00571] Taken together, these results establish that protein delivery of base editors maintains on-target base editing efficiency and greatly enhances editing specificity relative to delivery of plasmid DNA.

RNP delivery decouples on- and off-target editing

[00572] Given the striking enhancement of base editing specificity associated with protein delivery of BE3, it was investigated if this improvement was a result of a reduction in the total quantity of active genome editing agent delivered into the cell. Using the sgRNA targeting EMX1, a dose response study for plasmid (Fig. 128B) and protein delivery (Fig. 128C) was performed. To maximize transfection efficiency between treatment conditions, the volume of Lipofectamine 2000 was 1.5 μï¸ for all tests, and the base editor protein:sgRNA molar ratio was maintained at 1:1.1 for protein delivery. For plasmid delivery, a mass ratio of sgRNA

plasmid:BE3 plasmid of 1:3 (molar ratio ~1:1) and 1.5 μl of Lipofectamine 2000 were used. Off-target base editing was observed under all conditions tested for plasmid delivery (Fig. 128B), but virtually no off-target editing under all protein delivery conditions tested (Fig. 128C).

[00573] Linear regression analysis was performed to assess the relationship between on- and off-target editing for plasmid and protein delivery. For plasmid delivery, off-target editing was closely associated with on-target editing rates ($R^2 = 0.95$, $p = 0.0012$ for non-zero slope, F-test), whereas there was no significant association between off-target and on-target editing using protein delivery ($R^2 = 0.078$, $p = 0.59$ for non-zero slope, F-test).

[00574] These data indicate that protein delivery of base editors offers an inherent specificity advantage that is independent of dosage. Together with the previous observations[29,34], these findings support a model in which the higher DNA specificity of base editing from protein delivery compared to DNA delivery arises from the ability of protein delivery to avoid extended exposure of the genome to base editors, thereby minimizing the opportunity of base editors to process off-target loci after on-target loci have already been modified.

DNA-free base editing in zebrafish and mice

[00575] The above observations suggested the promise of protein delivery of BE3 to maintain on-target base editing while eliminating detectable off-target base editing. It was therefore tested whether protein delivery of BE3 could be used to generate specific point mutations in zebrafish by injecting BE3:sgRNA complexes targeting the *tyrosinase* locus into fertilized zebrafish embryos. Genomic DNA was harvested from the resultant zebrafish larvae 4 days post-injection and measured base editing and indel frequencies by high-throughput sequencing (Fig. 132A). Two of the three BE3:sgRNA complexes tested induced substantial point mutations *in vivo* (TYR1: $C_3 \rightarrow T_3$ 5.3+1.8%, TYR2: $C_4 \rightarrow T_4$ 4.3+2.1%; mean ± s.d. of n = 3 injected embryos; Fig. 132A). Sequences of zebrafish loci are listed in Table 17.

HEK cell samples

Sample Description	Replicate 1	2	3	Amplicon
Protein, BE3	237256	295609	159391	VEGFA On Target
Protein, HF-BE3	383480	389874	383467	VEGFA On Target
Plasmid, BE3	315213	280891	335668	VEGFA On Target
Plasmid, HF-BE3	196323	251965	369201	VEGFA On Target
Control	390748	395573	353614	VEGFA On Target
Protein, BE3	19280	26472	24799	FANCF Off Target Site #1
Protein, HF-BE3	36383	30007	39193	FANCF Off Target Site #1
Plasmid, BE3	35580	29557	22243	FANCF Off Target Site #1
Plasmid, HF-BE3	40371	27187	28248	FANCF Off Target Site #1
Control	35106	37274	34939	FANCF Off Target Site #1
Protein, BE3	82978	124689	83840	VEGFA Off Target Site #1
Protein, HF-BE3	142404	140482	142220	VEGFA Off Target Site #1
Plasmid, BE3	112027	117071	100889	VEGFA Off Target Site #1
Plasmid, HF-BE3	114187	98876	122553	VEGFA Off Target Site #1
Control	76854	90547	89271	VEGFA Off Target Site #1
Protein, BE3	14514	25513	19325	VEGFA Off Target Site #2
Protein, HF-BE3	24678	24363	25312	VEGFA Off Target Site #2
Plasmid, BE3	16945	19918	10225	VEGFA Off Target Site #2
Plasmid, HF-BE3	12200	14769	17797	VEGFA Off Target Site #2
Control	8739	13648	7818	VEGFA Off Target Site #2
Protein, BE3	179241	111693	173909	VEGFA Off Target Site #3
Protein, HF-BE3	243899	300503	276139	VEGFA Off Target Site #3
Plasmid, BE3	208476	291370	155430	VEGFA Off Target Site #3
Plasmid, HF-BE3	117174	154033	199152	VEGFA Off Target Site #3
Control	119263	170436	121686	VEGFA Off Target Site #3
Protein, BE3	237799	262947	185371	VEGFA Off Target Site #4
Protein, HF-BE3	313253	233699	244922	VEGFA Off Target Site #4
Plasmid, BE3	243094	230316	234421	VEGFA Off Target Site #4
Plasmid, HF-BE3	170958	160091	140693	VEGFA Off Target Site #4
Control	158691	148720	137270	VEGFA Off Target Site #4
Protein, BE3	28684	28237	43315	HEK3 On Target
Protein, HF-BE3	49300	42576	57690	HEK3 On Target
Plasmid, BE3	55008	55813	54310	HEK3 On Target
Plasmid, HF-BE3	55199	11384	7659	HEK3 On Target
Control	63741	42878	48524	HEK3 On Target
Protein, BE3	104822	181792	161090	HEK3 Off Target Site # 1
Protein, HF-BE3	204580	175561	177303	HEK3 Off Target Site # 1
Plasmid, BE3	178584	152264	206863	HEK3 Off Target Site # 1
Plasmid, HF-BE3	191297	138425	160789	HEK3 Off Target Site # 1
Control	190303	190061	190516	HEK3 Off Target Site # 1
Protein, BE3	146089	95113	135015	HEK3 Off Target Site # 2
Protein, HF-BE3	155947	136541	157991	HEK3 Off Target Site # 2
Plasmid, BE3	150036	128498	158905	HEK3 Off Target Site # 2
Plasmid, HF-BE3	371077	123642	142562	HEK3 Off Target Site # 2
Control	130322	134545	141833	HEK3 Off Target Site # 2
Protein, BE3	145058	175338	161837	HEK3 Off Target Site # 3
Protein, HF-BE3	212337	178993	179887	HEK3 Off Target Site # 3
Plasmid, BE3	186452	166500	80441	HEK3 Off Target Site # 3
Plasmid, HF-BE3	163732	118453	134719	HEK3 Off Target Site # 3
Control	131461	134470	155608	HEK3 Off Target Site # 3
Protein, BE3	41986	61678	67890	FANCF On Target
Protein, HF-BE3	41057	55850	86411	FANCF On Target
Plasmid, BE3	39114	48575	70074	FANCF On Target
Plasmid, HF-BE3	41617	56638	75718	FANCF On Target
Control	68852	59422	81265	FANCF On Target
Protein, BE3	113462	80529	191344	FANCF Off Target Site # 1
Protein, HF-BE3	202662	233981	203024	FANCF Off Target Site # 1
Plasmid, BE3	208912	202044	107234	FANCF Off Target Site # 1
Plasmid, HF-BE3	86494	113989	86807	FANCF Off Target Site # 1
Control	92255	72386	56661	FANCF Off Target Site # 1
Protein, BE3	96271	117442	84374	FANCF Off Target Site # 2
Protein, HF-BE3	105624	102312	105343	FANCF Off Target Site # 2
Plasmid, BE3	101002	98747	70052	FANCF Off Target Site # 2
Plasmid, HF-BE3	108966	69787	83184	FANCF Off Target Site # 2
Control	99986	100344	100659	FANCF Off Target Site # 2
Protein, BE3	25524	182451	65388	FANCF Off Target Site # 3
Protein, HF-BE3	71858	75553	71785	FANCF Off Target Site # 3
Plasmid, BE3	60980	57360	78169	FANCF Off Target Site # 3
Plasmid, HF-BE3	68316	34659	85718	FANCF Off Target Site # 3
Control	49685	57388	60418	FANCF Off Target Site # 3
Protein, BE3	46981	90793	79439	EMX1 On Target
Protein, HF-BE3	58629	71186	61575	EMX1 On Target
Plasmid, BE3	70817	82736	75706	EMX1 On Target
Plasmid, HF-BE3	77038	71123	78511	EMX1 On Target
Control	62183	48574	68439	EMX1 On Target
Protein, BE3	165905	257565	142888	EMX 1 Off Target Site # 1
Protein, HF-BE3	148339	151300	130712	EMX 1 Off Target Site # 1
Plasmid, BE3	101950	103226	203004	EMX 1 Off Target Site # 1
Plasmid, HF-BE3	167969	175193	97010	EMX 1 Off Target Site # 1
Control	101476	150435	102327	EMX 1 Off Target Site # 1
Protein, BE3	136738	213438	118711	EMX 1 Off Target Site # 2
Protein, HF-BE3	123413	126114	109375	EMX 1 Off Target Site # 2
Plasmid, BE3	85576	86600	169592	EMX 1 Off Target Site # 2
Plasmid, HF-BE3	140317	145738	137050	EMX 1 Off Target Site # 2
Control	84818	125139	85454	EMX 1 Off Target Site # 2
Protein, BE3	11940	36593	24946	EMX1 Off Target Site # 3
Protein, HF-BE3	26762	31566	36377	EMX1 Off Target Site # 3
Plasmid, BE3	32420	21547	14659	EMX1 Off Target Site # 3
Plasmid, HF-BE3	31427	16592	17385	EMX1 Off Target Site # 3
Control	17385	28128	32717	EMX1 Off Target Site # 3

Murine Samples from NIH 3T3 cell treatment

Sample Description	Replicate 1	2	3	Amplicon
Plasmid, BE3	16641	102216	46361	On Target VEGFA (Mus)
Plasmid, HF-BE3	89330	126545	100993	On Target VEGFA (Mus)
Protein, BE3	88998	81697	51124	On Target VEGFA (Mus)
Protein, HF-BE3	128218	29193	131515	On Target VEGFA (Mus)
Control	18767	38866	58985	On Target VEGFA (Mus)
Plasmid, BE3	174782	167504	182565	CFD Off Target 1
Plasmid, HF-BE3	167120	182520	192389	CFD Off Target 1
Protein, BE3	230569	212605	138144	CFD Off Target 1
Protein, HF-BE3	228668	211457	183370	CFD Off Target 1
Control	171738	191117	208791	CFD Off Target 1
Plasmid, BE3	206475	227332	206089	CFD Off Target 2
Plasmid, HF-BE3	213809	203028	199078	CFD Off Target 2
Protein, BE3	215995	275754	249969	CFD Off Target 2
Protein, HF-BE3	250918	272063	241059	CFD Off Target 2
Control	193760	175959	24696	CFD Off Target 2
Plasmid, BE3	60388	126278	7328	CFD Off Target 3
Plasmid, HF-BE3	89045	128508	5178	CFD Off Target 3
Protein, BE3	167195	330046	11163	CFD Off Target 3
Protein, HF-BE3	12120	309352	10393	CFD Off Target 3
Control	83204	176939	5661	CFD Off Target 3
Plasmid, BE3	192846	113709	171078	CFD Off Target 4
Plasmid, HF-BE3	205601	151434	188943	CFD Off Target 4
Protein, BE3	218194	181993	208398	CFD Off Target 4
Protein, HF-BE3	211966	148976	186838	CFD Off Target 4
Control	183933	130318	197476	CFD Off Target 4

Mouse Cochlea Samples

Sample Description	Replicate 1	2	3	Amplicon
Stria vascularies	37889	205706	62091	On Target (VEGFA)
Organ of Corti	148447	175004	29075	On Target (VEGFA)
Modiolus	182806	181382	61269	On Target (VEGFA)
Uninjected control	228222	241979	272759	On Target (VEGFA)
Stria vascularies	44457	244487	244646	CFD Off Target 1
Organ of Corti	136335	118318	34747	CFD Off Target 1
Modiolus	67176	209543	68699	CFD Off Target 1
Uninjected control	343100	342717	379015	CFD Off Target 1
Stria vascularies	72962	319883	265793	CFD Off Target 2
Organ of Corti	198456	131430	60530	CFD Off Target 2
Modiolus	92014	251509	81413	CFD Off Target 2
Uninjected control	399138	345965	483920	CFD Off Target 2
Stria vascularies	8325	80322	142556	CFD Off Target 3
Organ of Corti	81014	45976	1810	CFD Off Target 3
Modiolus	9928	75555	113411	CFD Off Target 3
Uninjected control	399138	345965	483920	CFD Off Target 3
Stria vascularies	232194	397770	554054	CFD Off Target 4
Organ of Corti	313472	285302	176872	CFD Off Target 4
Modiolus	230105	371399	258142	CFD Off Target 4
Uninjected control	524503	637946	624709	CFD Off Target 4

Zebrafish samples

	1	2	3	Amplicon
Treated zebrafish	72355	49498	81061	TYR 1
Scrambled sgRNA	107919	98502	92429	TYR 1
Treated zebrafish	51434	48014	41547	TYR 2
Scrambled sgRNA	61466	62374	66765	TYR 2
Treated zebrafish	6487	57247	75883	TYR 3
Scrambled sgRNA	64596	71234	75624	TYR 3

Mouse Cochlea Samples - treated with unrealated sgRNA

Sample			Amplicon
Stria	Cortil	Modiolus	
537459	249767	389274	On Target (VEGFA)

Table 17: Number of HTS reads that align to the reference sequence and pass the quality filters described in Materials and Methods.

[00576] Finally, these developments were applied to achieve DNA-free, high-specificity base editing in mice. To maximize the likelihood of observing on- and off-target base editing *in vivo,* the highly repetitive sgRNA targeting VEGFA site 2 was used; conveniently, the murine and human genomes are identical at this target site.

[00577] Using cultured murine NIH/3T3 cells, it was confirmed that BE3 protein delivery yielded efficient on-target base editing at this locus 34+11% (Fig. 133A; all editing percentages in this paragraph represent mean ± s.d. for n = 3 biological replicates). The Cutting Frequency Determinant (CFD) algorithm[29,34] was used to predict off-target loci in the mouse genome associated with the VEGFA site 2 sgRNA (Table 18). Using cultured NIH/3T3 cells, it was confirmed that two of the top four predicted off-target loci are indeed modified by plasmid delivery of BE3 in cultured murine cells (CFD off-target locus 1, 9+5% editing; and CFD off-target locus 4, 3+2% editing; Fig. 133B-133E). Consistent with the results from human cells, protein delivery of BE3 reduced off-target editing to levels similar to that of negative controls (Figs. 133C and 133E). The mean base editing specificity ratio for CFD off-target loci 1 and 4 increased from 28+13 for plasmid delivery of BE3 to ≥ 780+300 for protein delivery of BE3 (values represent mean ± s.e.m.; n = 3 biological replicates).

Site	Sequence	SEQ ID NO	CFD score	Description of locus
On-target	GACCCCCTCCACCCCGCCTCCGG	497		VEGFA site 2
Off-target 1	TCCCCCCTCCACCCCACCTCCGG	498	0.7857	intergenic:mmu-mir-21c-Nrp1/Mir1903
Off-target 2	TGCCCACCTCACCCCGCCTCTGG	499	0.65	intron:Vipr1
Off-target 3	GCCCCTCCCAACCCCACCTCTGG	500	0.6323	intron:Nos1ap
Off-target 4	CACCCCCCTCACCCCGCCTCAGG	501	0.625	intergenic:Unc5b-mmu-mir-6408

Table 18: Protospacer and PAM sequences for the predicted off-target loci in the mouse genome associated with the VEGFA site 2 sgRNA. CFD scores [46] were calculated using CRISPOR [47]. Positions in the off-target protospacers that differ from the on-target sequence are underlined.

[00578] To establish DNA-free base editing in mice, BE3:sgRNA complexes were combined with Lipofectamine 2000 (Fig. 132B) and intracochlear injections were performed into mouse pups at P1-P2. Injected cochlear tissues were harvested 3-4 days post-injection and micro-

dissected into 5-7 samples per cochlear region. Control cochlea from uninjected mice were harvested simultaneously. Genomic DNA was extracted from the harvested tissue, amplified by qPCR to late-exponential phase, and subjected to high-throughput DNA sequencing to measure C-^T conversion. Although it is impossible to quantitate base editing efficiency among treated cells because it is not possible to retrieve DNA exclusively from cells exposed to base editor protein, unambiguous base editing was observed from tissue in three regions of the cochlea: the basal end of the organ of Corti, the stria vascularis and the modiolus (Figs. 132C-132D). No significant indel formation was detected in treated tissue samples (< 0.1% indels; Fig. 134B).

[00579] The percentage of cochlear cells containing target C-^T conversion (Fig. 132C) was significantly lower than that observed in treated NIH/3T3 cells in culture (Fig. 133A), consistent with the highly localized nature of lipid-based protein delivery and the inability to isolate DNA exclusively from cells exposed to base editor. Nonetheless, local delivery offers key advantages for accessible applications, including control over which cell types are edited, and ease of preparation and administration.

[00580] Finally, off-target editing following intracochlear injection of BE3:sgRNA:lipid complexes was analyzed. Analysis of all four predicted off-target loci, including the confirmed off-target sites CFD locus 1 and CFD locus 4, in genomic DNA from the cochlear tissue of mice injected with the BE3:VEGFA site 2 sgRNA:lipid complex revealed no detectable C-^T conversion or indel formation above that observed in untreated controls samples for any of the off-target loci tested (Fig. 134A).

[00581] Together, these *in vivo* base editing results establish a virus-free, DNA-free strategy for the precise conversion of individual nucleotides in the genomic DNA of animals with high DNA sequence specificity.

Discussion

[00582] The strategies developed and implemented in this study expand the utility and applicability of base editing by removing or reducing off-target base editing and establishing a DNA-free delivery method that supports *in vivo* base editing. Protein delivery improves base editing specificity in human and murine cells compared with plasmid delivery of the same constructs (Figs. 137, 130, and 133), and enables specific base editing in zebrafish and in the mouse cochlea (Fig. 132).

[00583] A high-fidelity base editor was generated by installing into BE3 mutations known to enhance the DNA specificity of Cas9[20]. The installation of these mutations into Cas9 was reported to result in undetectable indel formation at off target loci associated with non-repetitive sgRNAs, including the EMX1 locus interrogated here (Fig. 127A)[20]. The specificity enhancements observed in HF-BE3, while substantial, were more modest; HF-BE3 exhibited detectable off-target base editing at both repetitive and non-repetitive loci when delivered as plasmid DNA into mammalian cells (Figs. 127A, 127D, 133C, and 133E). It is tempting to speculate that this specificity enhancement difference may arise from the fact that base editing, unlike Cas9-mediated indel formation, does not require DNA cleavage but only necessitates DNA-binding and R-loop formation[14], and some of the enhanced specificity of HF-Cas9 may arise from impaired DNA cleavage at already-bound off-target loci.

[00584] In a second attempt to reduce off-target base editing, it was demonstrated that RNP delivery of base editors leads to decoupling of on- and off- target editing (Fig. 128B-128C). RNP delivery ablated off-target editing at non-repetitive sites while maintaining on-target editing comparable to plasmid delivery (Fig. 130A-130C and 128A), and greatly reduced off-target editing even at the highly repetitive VEGFA site 2 (Fig. 130D). RNP delivery of base editors may be especially useful for *in vivo* editing applications in which cellular dosage is typically difficult to control or characterize.

[00585] RNP delivery of Cas9 coupled with delivery of a donor DNA template has previously been used to perform HDR-based genome editing in mammalian cells. These approaches, however, remain limited by low efficiency, cell-state dependence, and indel formation efficiencies typically exceeding those of desired HDR outcomes, especially for point mutation correction[29,30,32,35] DNA-free base editing, in contrast, generates a substantial excess of edited product relative to stochastic indels both *in vivo* and in cells (Fig. 132A, 134A, and 134B). To the best of the inventors' knowledge, RNP delivery of base editors represents the first strategy for generating specific and precise modifications to genomic DNA without requiring exogenous DNA.

Methods

Cloning ofplasmids

[00586] The plasmids in this study were generated by USER cloning. Phusion U Hot Start polymerase (Thermo Fisher) was used to install point mutations and construct protein expression plasmids from previously reported constructs [36]. Protein sequences are listed in the Supplementary Information, and plasmids for expression of BE3 and HF-BE3 are available from Addgene.

Expression and purification of BE3 and HF-BE3

[00587] BL21 Star (DE3)-competent *E. coli* cells were transformed with plasmids encoding the bacterial codon optimized base editors with a His_6 N-terminal purification tag. A single colony was grown overnight in Luria-Bertani (LB) broth containing 50 μg mL^{-1} kanamycin at 37 °C. The cells were diluted 1:200 into 2 L of the same media and grown at 37 °C until OD_{600} = 0.70-0.75. The cultures were incubated on ice for 60 min and protein expression was induced with 0.5 mM isopropyl -P-D-l-thiogalactopyranoside (IPTG, GoldBio). Expression was sustained for 14-16 h with shaking at 18 °C. The subsequent purification steps were carried out at 4 °C. Cells were collected by centrifugation at 6,000 *g* for 20 min and resuspended in cell collection buffer (100 mM tris(hydroxymethyl)-aminomethane (Tris)-HCl, pH 8.0, 1 M NaCl, 20% glycerol, 5 mM tris(2-carboxyethyl)phosphine (TCEP; GoldBio), 0.4 mM phenylmethane sulfonyl fluoride (PMSF; Sigma Aldrich) and 1 cOmplete, EDTA-free protease inhibitor pellet (Roche) per 50 mL buffer used). Cells were lysed by sonication (6 min total, 3 s on, 3 s off) and the lysate cleared by centrifugation at 25,000 *g* (20 min).

[00588] The cleared lysate was incubated with His-Pur nickel nitriloacetic acid (nickel-NT A) resin (1 mL resin per litre of culture, Thermo Fisher) with rotation at 4 °C for 60-90 min. The resin was washed with 20 column volumes of cell collection buffer before bound protein was eluted with elution buffer ((100 mM tris(hydroxymethyl)-aminomethane (Tris)-HCl, pH 8.0, 0.5 M NaCl, 20% glycerol, 5 mM tris (2-carboxyethyl) phosphine (TCEP; GoldBio), 200 mM imidazole). The resulting protein fraction was further purified on a 5 mL Hi-Trap HP SP (GE Healthcare) cation exchange column using an Akta Pure FPLC. Protein-containing fractions were concentrated using a column with a 100,000 kDa cutoff (Millipore) centrifuged at 3,000 *g* and the concentrated solution was sterile filtered through an .22 μιη PVDF membrane (Millipore).

[00589] After sterile filtration, proteins were quantified with Reducing Agent Compatible Bicinchoninic acid (BCA) assay (Pierce Biotechnology), snap-frozen in liquid nitrogen and stored in aliquots at -80 °C. Sequences of expressed proteins are listed in Supplementary Note 2.

In vitro transcription ofsgRNA

[00590] Linear DNA fragments containing the T7 RNA polymerase promoter sequence upstream of the desired 20 bp sgRNA protospacer and the sgRNA backbone were generated by PCR (Q5 Hot Start MasterMix, New England Biolabs) using primers as listed in the Supplementary Information and concentrated on minelute columns (Qiagen). sgRNA was transcribed with the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) at 16 °C for 14-16 h with 1 μg of linear template per 20 μ reaction. sgRNA was purified using the MEGAClear Transcription Clean Up Kit (Thermo Fisher), according to the manufacturer's instructions. Purified sgRNAs were stored in aliquots at -80 °C.

In vitro deamination assays

[00591] Sequences of DNA oligonucleotides used as templates for the *in vitro* deamination assay are shown in Supplementary Note 3. All oligonucleotides were purchased from IDT. Single-stranded oligonucleotides synthesized with complementary sequences were combined (5 μL of a 100 μM solution) in Tris buffer pH 8.0 and annealed by heating to 95 °C for 5 min, followed by a gradual cooling to 37 °C at a rate of 0.1 °C $second^{-1}$ to generate 79 base pair (bp) dsDNA substrates. Freshly thawed base-editor proteins (2 μM final concentration in a 10 μL reaction volume) were complexed with the indicated sgRNA (2.2 μM final concentration) in Reaction Buffer (20 mM HEPES pH 7.5, 150 mM KC1, 0.5 mM DTT, 0.1 mM EDTA, 10 mM $MgCl_2$)[37] for five minutes at room temperature. Annealed dsDNA substrates were then added to a final concentration of 250 nM. The reaction proceeded for 30 min at 37 °C before protein denaturation was performed by heating for 5 min at 99 °C. Addition of PB buffer (Qiagen, 100 μL) and isopropanol (25 μL) ensured protein was dissociated from the substrate DNA. DNA was purified with Minelute columns (Qiagen) and the resulting products amplified to the top of the linear range with 15 cycles of qPCR (12 ng input DNA, 50 μL reaction volume) using a U-tolerant polymerase (Phusion U Hot Start, ThermoFisher) and primers as listed in the Supplementary Information. Amplified DNA was purified using RapidTip2 (Diffinity Genomics)

and barcoded with a second round of PCR (8 cycles, 5 ng input) before being prepared for sequencing on an Illumina MiSeq as described below.

Purification and sequencing of genomic DNA

[00592] Genomic DNA was isolated using Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) according to the manufacturer's instructions. For the first PCR, DNA was amplified to the top of the linear range using Q5 Hot Start DNA Polymerase (NEB), according to the manufacturer's instructions but with the addition of 3% DMSO and SYBR Gold Nucleic Acid Stain (Thermo Fisher). For all amplicons, the PCR protocol used was an initial heating step of 2 min at 98 °C followed by an optimized number of amplification cycles (12 s at 98 °C, 25 s at 61 °C, 30 s at 72 °C). For zebrafish and for transfected cell samples, 30 ng of input DNA was used in a 50 μl reaction, for cochlear samples 20 ng was used in a 25 μl reaction. qPCR was performed to determine the optimal cycle number for each amplicon. Amplified DNA was purified using RapidTip2 (Diffinity Genomics) and barcoded with a further PCR (8 cycles, 5 ng input). The unique forward and reverse primers used in the first-round PCR contained a constant region 5' to the annealing region, (forward: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN-3' (SEQ ID NO: 502), reverse: 5'-TGGAGTTCAGACGTGTGCTCTTCCGATCT-3'(SEQ ID NO: 503)) which facilitated binding of barcoding primers to amplified DNA for a second-round PCR.

[00593] The second-round PCR used primers with three regions: a 5' constant region allowing the amplicon to bind to the Illumina flow cell *(italicized)*, an 8-base barcoding region (X), and a 3' constant region allowing the barcoding primer to bind to the first-round PCR amplicon (in **bold).** Examples of primer sequences are:

forward: 5'-*AATGATACGGCGACCACCGAGATCTACACXXXXXXXXACACTCTTTCCCTACACGAC***-3'** (SEQ ID NO: 504)

reverse: 5'-CAAGCAGAAGACGGCATACGAGATXXXXXXXGTGACTGGAGTTCAGACGTGTGC TCTTC **-3'** (SEQ ID NO: 505)

[00594] Sequencing adapters and dual-barcoding sequences are based on the TruSeq Indexing Adapters (Illumina). Barcoded samples were pooled and purified by gel extraction (Qiagen), and

then purified using Ampure beads (Beckman Coulter) before quantification using the Qubit dsDNA HS Kit (Thermo Fisher) and qPCR (KAPA BioSystems) according to the manufacturer's instructions. Sequencing of pooled samples was performed using a single-end read from 180-250 bases (depending on the amplicon size) on the MiSeq (Illumina) according to the manufacturer's instructions.

[00595] Sequences of oligonucleotides used for PCR amplification are shown in Supplementary Note 3. All oligonucleotides were obtained from IDT. The optimized number of PCR cycles for each amplicon in this study are as follows: VEGFA site 2 human genomic DNA (annealing temperature was 61 °C for 25 seconds for all extension steps): on-target: 29 cycles, off-target #1: 32 cycles, off-target #2: 28 cycles, off-target #3: 27 cycles, off-target #4: 27 cycles, VEGFA site 2 murine genomic DNA: on-target: 31 cycles, off-targets #1, #2, #3 and #4: 31 cycles. HEK293 site 3: off-targets #1: 29 cycles, off-target #2: 28 cycles, off-target #3: 28 cycles. FANCF off-target #1: 29 cycles, off-target #2: 28 cycles, off-target 3: 28 cycles. EMX1 off-targets #1, #2 and #3: 28 cycles. TYR1, TYR2 and TYR3 sgRNAs for amplification of zebrafish DNA: 32 cycles. Optimized protocols for the on-target amplification of the EMX1, FANCF, and HEK293 site 3 loci were followed as previously described [14].

Analysis and alignment of genomic DNA sequencing reads

[00596] Sequencing reads were analyzed as previously described [14]. In brief, sequencing reads were demultiplexed using MiSeq Reporter (Illumina), and individual FASTQ files were analyzed with a previously reported custom Matlab script [14]. Reads were aligned to the reference sequence using the Smith-Waterman algorithm. Base calls with Q-scores below 30 were replaced with a placeholder nucleotide (N). This quality threshold results in nucleotide frequencies with an expected error rate of 1 in 1,000. Indel frequencies were quantified with a previously published custom Matlab script which counts indels which occurring in a 30-base window around the nCas9 cleavage site and are a minimum of 2-base insertions or deletions [14]. Indels were defined as detectable if there was a significant difference (Student's two-tailed t-test, $p < 0.05$) between indel formation in the treated sample and untreated control.

[00597] For one of the sequenced amplicons, CFD off-target #3, associated with VEGFA site 2 sgRNA in the murine genome, it was not possible to accurately measure indel formation. The protospacer at this locus is directly preceded by 12 guanine bases, which makes PCR and high-

throughput sequencing of this site prone to random insertion or deletions; deletion rates as high as 20% of sequencing reads were observed in multiple independent untreated control samples. Since no significant base editing was detected at this off-target locus under any treatment conditions (Fig. 132 andl33), it is suspected that indel formation is also negligible at this locus.

[00598] A phred.II Q30 score corresponds to an estimated 99.9% accuracy in basecalling[38]. A 0.1% probability of incorrect base calling at a given position corresponds to a lower limit for base calling of 0.1/4 = 0.025% if it is assumed base call errors are randomly distributed across the four bases. C-^T editing percentages that fell beneath this threshold were classified as undetectable. Spontaneous deamination[39] or polymerase error during PCR can also introduce artefactual C-^T edits. In order to distinguish base editor-induced C-^T editing from artefactual C-^T editing rates, untreated control cells were sequenced for each amplicon and it was calculated whether the C-^T editing under a particular condition was statistically significant using the Student's two-tailed t-test with $p < 0.05$ as the threshold. Off-target sites with statistically significant editing rates >0.025% were considered measureable. The number of aligned and quality filtered reads for each sample has been included in Table 17.

Statistical analyses of genomic DNA sequence alignments

[00599] Unless otherwise noted, mean values cited throughout the main text are representative of n>3 independent biological replicates and the mean ± standard deviation has been stated.

[00600] The statistical analysis of the high-throughput sequencing data displayed in Figs. 2 and 3 was performed by comparing on- and off-target editing percentages in treated samples to any editing measured in a negative control sample (untreated). The Student's two-tailed t test was used, and individual /^-values are shown in Table 16. $*p \leq 0.05$, $**p \leq 0.01$ and $***p \leq 0.001$. When editing was below the detection limit (0.025%), significance was not calculated; all untreated control samples showed undetectable editing.

[00601] For Fig. 128A, mean on-target base editing was calculated by averaging editing of cytosines in the base editing activity window (C_4-C_8 for HEK293 site 3 and EMX1, C_4-C_9 for FANCF and VEGFA site 2).

[00602] To account for sgRNA-dependent differences in base editing activity, the a base editingdndel ratio was calculated (Fig. 130B). This ratio was generated by dividing the percentage of HTS reads with a C-^T conversion (averaged across the base editing window for

each site) by the percentage of HTS reads containing an indel. As described above, if the off-target editing for a particular locus was below the limit of detection it was conservatively assumed the estimated upper bound of the detection method (0.025%) for the purpose of calculating specificity ratios.

Data analysis of in vitro edited DNA

[00603] Sequencing reads were automatically demultiplexed using MiSeq Reporter (Illumina.). Quality filtering was performed using the online package usegalaxy.org[40]. Individual bases with an Illumina quality score less than or equal to 30 were converted to the placeholder nucleotide 'N' using FASTQ Groomer followed by FASTA Masker[41]. The resulting quality-filtered FASTQ files were subsequently analysed with a custom python script provided in Supplementary Note 1. Sequencing reads were scanned for exact matches to two 14-base sequences that flank both sides of the target DNA sequence. If no exact matches were found, the read was excluded from analysis. If both 14-base sequences were located and the length of the sequence between them was equal to the expected protospacer length (20 bases), the protospacer sequence found between the flanking regions was saved and the bases called by high-throughput sequencing at each site within the protospacer were tallied.

Cell culture

[00604] Both HEK293T (ATCC CRL-3216) and NIH/3T3 (ATCC CRL-1658) were maintained in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher) supplemented with 10 % (v/v) fetal bovine serum (FBS), at 37 °C with 5 % $C0_2$. Cells were obtained from ATCC and were authenticated and verified to be free of mycoplasma by ATCC upon purchase.

Plasmid transfection of base editors into HEK293T cells

[00605] HEK293T cells were seeded on 48-well collagen-coated BioCoat plates (Corning) in antibiotic free medium and transfected at approximately 70 % confluency. Unless otherwise noted, 750 ng of BE and 250 ng of sgRNA expression plasmids were transfected using 1.5 μl of Lipofectamine 2000 (Thermo Fisher) per well according to the manufacturer's protocol.

Protein transfection of base editors into HEK293T cells

[00606] HEK293T cells were seeded on 48-well collagen-coated BioCoat plates (Corning) in 250 μï¸ antibiotic free medium and transfected at approximately 70 % confluency. Base editor protein and was incubated with 1.1X molar excess of the necessary sgRNA at room temperature for 5 min. The complex was then incubated with 1.5 μï¸ Lipofectamine 2000 (Thermo Fisher) and transfected according to the manufacturer's protocol for plasmid delivery. Unless otherwise noted, BE protein was added to a final concentration of 200 nM (based on a total well volume of 275 μL).

Plasmid transfection of base editors into NIH/3T3 cells

[00607] NIH/3T3 cells were seeded on 48-well collagen-coated BioCoat plates (Corning) in antibiotic-free DMEM medium and transfected at approximately 75 % confluency. Unless otherwise noted, 600 ng of BE and 200 ng of sgRNA expression plasmids were transfected using 1.4 μL of Lipofectamine 3000 with 1 μL of P3000 reagent (Thermo Fisher) per well according to the manufacturer's protocol.

Protein transfection of base editors into NIH/3T3 cells

[00608] NIH/3T3 cells were seeded on 48-well collagen-coated BioCoat plates (Corning) in antibiotic free DMEM medium and transfected at approximately 75 % confluency. Base editor proteins were incubated with 1.1-fold molar excess of the indicated sgRNA at 25 °C for 5 min. The complex was then incubated with 1.4 μL Lipofectamine 3000 (Thermo Fisher) and transfected according to the manufacturer's protocol for plasmid delivery. P3000 reagent was not used because its addition lead to protein precipitation and a reduction in base editing efficiency. Unless otherwise noted, BE protein was added to a final concentration of 400 nM (based on a total well volume of 275 μL).

Intracochlear delivery ofBE3 protein:guide RNA encapsulated in cationic lipid

[00609] All animal experiments were approved by the Institutional Animal Care and the Use Committee of the Massachusetts Eye and Ear Infirmary. Intracochlear delivery was performed in P1-P2 mice of a mixed genetic background as described previously[42]. Mice were anesthetized by lowering body temperature before the surgical procedure. A postauricular incision was made

near the right ear, and the bulla was lifted to expose the cochlea. BE3 protein (57.7 μM) was pre-complexed with the sgRNA (100 μM) in a 1:1.1 molar ratio and then mixed with Lipofectamine 2000 (Thermo Fisher) in a 1:1 volumetric ratio. The resulting solution (1.2-1.5 μL) was injected with a glass pipette (end diameter, 5 μιη) through the cochlear capsule into scala media at the cochear basal turn that attached to a nanoliter micropump (WPI, UMP3 + Micro4 + NanoFil) at the rate of 250 nL min^{-1}. After injection, the incision was closed and the mice were brought onto a heating pad to recover. After 3-4 days, the cochlea of mouse was dissected into the organ of Corti, stria vascularis, and modiolus. Each tissue was further micro-dissected into between 5 and 7 separate pieces and DNA extraction was performed separately for each sample, followed by high-throughput sequencing as described above. The data presented in Fig. 132 and Fig. 134 show sequencing data resulting from extraction of one micro-dissected sample for each cochlear region.

Microinjection ofBE3 protein:guide RNA into zebrafish embryo

[00610] Zebrafish (Tuebingen strain) were maintained under standard conditions in compliance with internal regulatory review at Boston Children's Hospital. One-cell stage zebrafish embryos were injected with approximately 2 nL of BE3 protein pre-complexed with the appropriate sgRNA or an unrelated sgRNA control in a 1:1 molar ratio (4.5 μM final concentration). Four days post-fertilization, DNA was extracted from larvae as previously described[43] in 50 mM NaOH for 30 minutes at 95 °C and the resulting solution was neutralized with Tris-HCl. Genomic DNA was quantified, amplified by PCR, and sequenced as described above.

Protein gel analyses

[00611] All protein gels shown were precast 4-12% polyacrylamide Bis-Tris Plus (Thermo Fisher). They were run in MOPS buffer (Thermo Fisher) at 180 V for 50 min. Samples were prepared for loading by heating to 99 °C in lOOmM DTT and IX lithium dodecyl sulfate (LDS) Sample Buffer for denaturation (Thermo Fisher) for 10 min. Gels were stained using Instant Blue Protein Stain (Expedion) according to manufacturer's instructions.

[00612] For cell lysate analysis, 2 niL of post-induction overnight culture was pelleted at 15,000 *g* before lysis in 100 μï¸ B-PER (Thermo Fisher) according to the manufacturer's instructions.

Data Availability

[00613] High-throughput sequencing data that support the findings of this study have been deposited in the NCBI Sequence Read Archive database under Accession Number SRP097884. Plasmids encoding HF-BE3 and BE3 for protein expression, as well as HF-BE3 for mammalian expression, are available from Addgene with Accession IDs 87439 (pCMV~HF-BE3). 87438 (pET42b-HF-BE3), 87437 (pET42b-BE3).

References

1 Cox, D. B., Piatt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nature medicine* **21,** 121-131, doi:10.1038/nm.3793 (2015).
2 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346,** 1258096, doi: 10. 1126/science. 1258096 (2014).
3 Komor, A. C , Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168,** 20-36, doi:10.1016/j.cell.2016.10.044 (2017).
4 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32,** 347-355, doi:10.1038/nbt.2842 (2014).
5 Hilton, I. B. & Gersbach, C. A. Enabling functional genomics with genome engineering. *Genome Res* **25,** 1442-1455, doi: 10. 1101/gr. 190124. 115 (2015).
6 Wyman, C. & Kanaar, R. DNA double-strand break repair: all's well that ends well. *Annual review of genetics* **40,** 363-383, doi: 10. 1146/annurev.genet.40. 110405.090451 (2006).
7 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8,** 2281-2308, doi:10.1038/nprot.2013.143 (2013).
8 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819-823, doi: 10. 1126/science. 123 1143 (2013).
9 Bodles-Brakhop, A. M., Heller, R. & Draghia-Akli, R. Electroporation for the Delivery of DNA-based Vaccines and Immunotherapeutics: Current Clinical Developments. *Mol Ther* **17,** 585-592, doi:10.1038/mt.2009.5 (2009).
10 Kay, M. A., Glorioso, J. C. & Naldini, L. Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nature medicine* **7,** 33-40, doi:Doi 10.1038/83324 (2001).
11 Midoux, P., Pichon, C , Yaouanc, J. J. & Jaffres, P. A. Chemical vectors for gene delivery: a current review on polymers, peptides and lipids containing histidine or imidazole as nucleic acids carriers. *Brit J Pharmacol* **157,** 166-178, doi:10.1111/j.l476-5381.2009.00288.x (2009).
12 Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nature biotechnology* **32,** 551-553, doi:10.1038/nbt0914-952d (2014).

13 Hu, J. H., Davis, K. M. & Liu, D. R. Chemical Biology Approaches to Genome Editing: Understanding, Controlling, and Delivering Programmable Nucleases. *Cell chemical biology* **23,** 57-73, doi:10.1016/j.chembiol.2015. 12.009 (2016).
14 Komor, A. C , Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double- stranded DNA cleavage. *Nature* **533,** 420-424, doi:10.1038/naturel7946 (2016).
15 Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nature methods* **13,** 1036-1042, doi:10.1038/nmeth.4038 (2016).
16 Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nature methods* **13,** 1029-1035, doi: 10.1038/nmeth. 4027 (2016).
17 Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353,** aaf8729, doi:10.1126/science.aaf8729 (2016).
18 Li, J., Sun, Y., Du, J., Zhao, Y. & Xia, L. Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol Plant* **In Press,** doi:10.1016/j.molp.2016.12.001 (2016).
19 Koyama, S., Ishii, K. J., Coban, C. & Akira, S. Innate immune response to viral infection. *Cytokine* **43,** 336-341, doi:10.1016/j.cyto.2008.07.009 (2008).
20 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529,** 490-495, doi:10.1038/naturel6526 (2016).
21 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31,** 827-832, doi:10.1038/nbt.2647 (2013).
22 Fu, Y. F. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31,** 822-826, doi:10.1038/nbt.2623 (2013).
23 Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology* **31,** 839-843, doi:10.1038/nbt.2673 (2013).
24 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology* **33,** 187-197, doi:10.1038/nbt.3117 (2015).
25 Guilinger, J. P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nature methods* **11,** 429-435, doi:10.1038/nmeth.2845 (2014).
26 Singh, R., Kuscu, C , Quinlan, A., Qi, Y. J. & Adli, M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic acids research* **43,** 677-683, doi:ARTN e l 18
10.1093/nar/gkv575 (2015).
27 Liang, X. Q. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* **208,** 44-53, doi:10.1016/j.jbiotec.2015.04.024 (2015).
28 Wang, M. *et al.* Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *Proceedings of the National Academy of Sciences of the United States of America* **113,** 2868-2873, doi: 10. 1073/pnas. 15202441 13 (2016).
29 Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature biotechnology* **33,** 73-80, doi:10.1038/nbt.3081 (2015).

30 DeWitt, M. A. *et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci Transl Med* **8,** 360ral34, doi:10.1126/scitranslmed.aaf9336 (2016).
31 Kim, K. *et al.* Highly efficient RNA-guided base editing in mouse embryos. *Nature biotechnology,* doi:10.1038/nbt.3816 (2017).
32 Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology* **34,** 339-344, doi:10.1038/nbt.3481 (2016).
33 Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature communications* **5,** 4767, doi:10.1038/ncomms5767 (2014).
34 Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* **11,** 316-318, doi:10.1038/Nchembio.l793 (2015).
35 Lin, S., Staahl, B., Alia, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife,* e04766, doi:10.7554/eLife.04766 (2014).
36 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology* **34,** 184-191, doi:10.1038/nbt.3437 (2016).
37 Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816-821, doi: 10. 1126/science. 1225829 (2012).
38 Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8,** 186-194 (1998).
39 Liu, M. & Schatz, D. G. Balancing AID and DNA repair during somatic hypermutation. *Trends in immunology* **30,** 173-181, doi:10.1016/j.it.2009.01.007 (2009).
40 Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research* **44,** W3-W10, doi:10.1093/nar/gkw343 (2016).
41 Blankenberg, D. *et al.* Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26,** 1783-1785, doi:10.1093/bioinformatics/btq281 (2010).
42 Hu, L. X. *et al.* Diphtheria Toxin-Induced Cell Death Triggers Wnt-Dependent Hair Cell Regeneration in Neonatal Mice. *J Neurosci* **36,** 9479-9489, doi:10.1523/Jneurosci.2447-15.2016 (2016).
43 Meeker, N. D., Hutchinson, S. A., Ho, L. & Trede, N. S. Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *Biotechniques* **43,** 610, 612, 614 (2007).
44 Zuris, J.A. et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nature biotechnology 33, 73-80 (2015).
45 Tsai, S.Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature biotechnology 33, 187-197 (2015).
46 Doench, J.G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature biotechnology 34, 184-191 (2016).
47 Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol 17, 148 (2016).

Supplementary Information

Supplementary Note 1: **Python script used to analyze quality-filtered *in vit*ro-edited DNA.**

```
1.      from __future__ import print_function
2.      from __future__ import division
3.
4.      import Bio #This will import the BioPython suite
5.      from Bio import SeqIO #Necessary to read/write sequence handles
6.      from Bio.Seq import Seq
7.      import os
8.      import collections
9.      import csv
10.
11.     inputfile = "please_specify_your_input_file_here_containing_filtered_reads" #specify the filenames that contain sequences
12.     filenames = []
13.
14.     for file in os.listdir(inputfile):
15.         if file.endswith(".fastqsanger") :
16.             filenames.append(file)
17.
18.     spacer = []
19.     list_of_filenames = []
20.
21.     for file in filenames:
22.         site = {}
23.         output = open(file + ".txt", "w")
24.         list_of_filenames.append(file + ".txt") #allows calling of the txt files that come from fastq files later
25.         for rec in SeqIO.parse(file, "fastq"):
26.             splitl=rec.seq.tostring().split("GTTCGCGGCGATCG")   #14-base pair constant_region_before_protospacer
27.             if len(splitl)>=2:
28.                 split2=splitl[1].split("TGGATCGCCTGGCA") #14-base pairc constant_region_after_protospacer
29.                 site=split2[0]
30.                 if len(site)==20:
31.                     output.write(site + "\n")
32.
33.     BASES = 'ATGCN'
34.     UNRECOGNIZED = 'X'
35.     BASE_SEPERATOR = dict(zip(BASES, ',,,,\n'))
36.     a_index = 0
37.     t_index = 1
38.     g_index = 2
39.     c_index = 3
40.     n_index = 4
41.
42.     def get_counts_by_column(base, count, library):
43.         current_count = library[count]
44.         if base == 'A':
45.             current_count[a_index] += 1
46.         elif base == 'T':
47.             current_count[t_index] += 1
48.         elif base == 'G':
49.             current_count[g_index] += 1
50.         elif base == 'C':
51.             current_count[c_index] += 1
52.         elif base == 'N':
53.             current_count[n_index] += 1
54.
```

```
    def dna_counts(list_of_sequences,  sample):
        first_oligo  = list_of_sequences[0]
        for i in range  (len(first_oligo)) :
            sample. append ([0,0, 0,0,0])
        for j in range(len(f irst_oligo) ):
            for  i in range(len(list_of_sequences)  ):
                get_counts_by_column(list_of_sequences[i]  [j], j, libname)


    for file in list_of_f ilenames :
        spacer_list  = open(f ile) .read( ).splitlines( )
        output2=[]
        dna_counts(spacer_list,  output2)
        with open(file  + ".csv",  "wb") as f :
            writer = csv.writer(f)
            writer.writerows(output2)
```

Supplementary Note 2: **Sequences of proteins used in this study**

Protein sequence of expressed BE3

MGSSHHHHHHSSETGPVAVDPTLRRRI EPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFI EKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGL
RDLISSGVTIQIMTEQESGYCW RNFVNYSPSN EAHWPRYPHLWVRLYVLELYCI ILGLPPCLNILRRKQPQ
LTFFTIALQSCHYQRLPPH ILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSK
KFKVLGNTDRHSIKKNLIGALLFDSG ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKH ERH PI FGNIVDEVAYH EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLI EG DL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG EKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYI DGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKI EKILTFRIPYYVG PLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFI ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSG EQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKI IKDK
DFLDNEENEDILEDIVLTLTLFEDREMI EERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING IRDKQS
GKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKG ILQTVKVVDEL
VKVMG RHKPENIVI EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH PVENTQLQNEKLYLYYLQ
NGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSI DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANG EI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GG FSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG ITIMERSS
FEKN PI DFLEAKGYKEVKKDLI IKLPKYSLFELENG RKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKH RDKPI REQAENIIHLFTLTNL
GAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG DSGGSTNLSDI IEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNG ENKIKMLSGGSPKK
KRKV (SEQ ID NO: 185)

Protein sequence of expressed HF-BE3

MGSSHHHHHHSSETGPVAVDPTLRRRI EPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFI EKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGL
RDLISSGVTIQIMTEQESGYCW RNFVNYSPSN EAHWPRYPHLWVRLYVLELYCI ILGLPPCLNILRRKQPQ
LTFFTIALQSCHYQRLPPH ILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSK
KFKVLGNTDRHSIKKNLIGALLFDSG ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL

EESFLVEEDKKH ERH PIFGNIVDEVAYH EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLI EG DL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG EKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYI DGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKI EKILTFRIPYYVG PLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFI ERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSG EQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKI IKDK
DFLDNEENEDILEDIVLTLTLFEDREMI EERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGI RDKQS
GKTILDFLKSDG FANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLH EHIANLAGSPAIKKGILQTVKVVDEL
VKVMG RHKPENIVI EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH PVENTQLQNEKLYLYYLQ
NGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSI DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLI REVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANG EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GG FSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG ITIMERSS
FEKN PIDFLEAKGYKEVKKDLI IKLPKYSLFELENG RKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKH RDKPI REQAENIIHLFTLTNL
GAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG DSGGSTNLSDI IEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNG ENKIKMLSGGSPKK
KRKV (SEQ ID NO: 186)

Supplementary Note 3: **Sequences of oligonucleotides used in the present study**

Unpublished Primers used to amplify off target genomic DNA for HTS in human cells (SEQ ID NOS: 506-513)

fwd_VEGFA_site2_off_target_1_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTACAAGTAACAGTCCAAGAA
rev_VEGFA_site2_off_target_1_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCTGCAACTTAACTTACGTGAAA
fwd_VEGFA_site2_off_target_2_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCAAGCCCATTTGTCCAGG
rev_VEGFA_site2_off_target_2_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTCTTTTTGAGCTTTGGGC
fwd_VEGFA_site2_off_target_3_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCATACCAGCAGCAGTTCC
rev_VEGFA_site2_off_target_3_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCACCTCAGCTCCTGCAC
fwd_VEGFA_site2_off_target_4_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCACTGATTCTACACCATGGT
rev_VEGFA_site2_off_target_4_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGTTCCCAACCTTTTTGACA

Other primers (for off target sites associated with HEK_3, EMX1, FANCF) were previously published

Primers used to amplify off target genomic DNA for HTS in murine cells (SEQ ID NOS: 514-521)

fwd_VEGFA_site2_off_target_1_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGCTGGAGATTCAGAGACAC
rev_VEGFA_site2_off_target_1_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGCCCCTTCTGACACACATAC
fwd_VEGFA_site2_off_target_2_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCCCTCAAGGCTTGACATTTC
rev_VEGFA_site2_off_target_2_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAAAAGTTGGGAGAGGGGATG
fwd_VEGFA_site2_off_target_3_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGTACCCCAGTCCCCTCATC
rev_VEGFA_site2_off_target_3_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAAGTTACGGGGATGTCACTTG
fwd_VEGFA_site2_off_target_4_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTAACATCCAGTCTCCCAAACACA
rev_VEGFA_site2_off_target_4_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTACACACACACACTACTAGGACA

Primers used to amplify on target genomic DNA for HTS in murine cells (SEQ ID NOS: 522-523)

fwd_VEGFA_site2_on_target_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGCTACTACGGAGCGAGAAG
rev_VEGFA_site2_on_target_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTACAGGGGCAAAGTGAGTGAC

Primers used for generating PCR products to serve as substrates for T7 transcription of sgRNAs (SEQ ID NOS: 524-529)

rev_sgRNA_T7: used in all cases	AAAAAAAGCACCGACTCGGTGCCAC
fwd_sgRNA_T7_EMX1	TAATACGACTCACTATAGGGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_FANCF	TAATACGACTCACTATAGGGGAATCCCTTCTGCAGCACCGTTTTAGAGCTAGAAATAGCA

fwd_sgRNA_T7_HEK_site_3	TAATACGACTCACTATAGGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_VEGFA_site_2	TAATACGACTCACTATAG GACCCCCTCCACCCCGCCTCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_TC_repeat_in_vitro	TAATACGACTCACTATAGGTCTCTCTCTCTCTCTCTCTCGTTTTAGAGCTAGAAATAGCA

Primers used for generating sgRNA transfection plasmids (SEQ ID NOS: 530-531)

The pFYF1320 plasmid was used as template as previously described (Komor et al). The sequence of other sgRNA plasmids was previously reported

rev_sgRNA_plasmid	GGTGTTTCGTCCTTTCCACAAG
fwd_VEGFA_site_2	GACCCCCTCCACCCCGCCTCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC

Seguences of ssDNA substrates used in in vitro deaminase assays (SEQ ID NOS: 532-533)

fwd_TC_repeat_substrate	ACGTAAACGGCCACAAGTTCGCGGCGATCGTCTCTCTCTCTCTCTCTCTCTGGATCGCCTGGCATCTTC TTCAAGGACG
rev_TC_repeat_substrate	CGTCCTTGAAGAAGATGCCAGGCGATCCAGAGAGAGAGAGAGAGAGAGACGATCGCCGCGAACTTGT GGCCGTTTACGT

Previously published primers used to amplify off target genomic DNA for HTS in human cells (SEQ ID NOS: 534-557)

fwd_EMX1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
rev_EMX1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
fwd_FANCF_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATTGCAGAGAGGCGTATCA
rev_FANCF_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
fwd_HEK293_site3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG
rev_HEK293_site3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC
fwd_EMX1_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGTAGCCTCTTTCTCAATGTGC
rev_EMX1_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTCACAAGGATGCAGTCT
fwd_EMX1_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTAGACTCCGAGGGGA
rev_EMX1_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTCGTCCTGCTCTCACTT
fwd_EMX1_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGAGGCTGAAGAGGAAGACCA
rev_EMX1_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCCAGCTGTGCATTCTAT
fwd_FANCF_off 1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACCCACTGAAGAAGCAGGG
rev_FANCF_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCTTAATCCGGCTCCAT
fwd_FANCF_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNTCCAGTGTTTCCATCCCG AA
rev_FANCF_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTGACCTCCACAACTCT

fwd_FANCF_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGGTACAGTTCTGCGTGT
rev_FANCF_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCACTCTGAGCATCGCCAAG
fwd_HEK293_site3_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCCCTGTTGACCTGGAGAA
rev_HEK293_site3_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCCCTGACCA
fwd_HEK293_site3_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGGTGTTGACAGGGAGCAA
rev_HEK293_site3_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGCAGAAGGG
fwd_HEK293_site3_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGAGGGAACAGAAGGGCT
rev_HEK293_site3_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCCAAGAACCT

Primers used to amplify on target genomic DNA for HTS in zebrafish (SEQ ID NOS: 558-563)

fwd_TYR1_zebrafish	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTTCCCCCGAGTCTGCACCT
rev_TYR1_zebrafish	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGAACTTGCATTCGCCGCAA
fwd_TYR2_zebrafish	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTCTGCCTTGGCATCGGGTG
rev_TYR2_zebrafish	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCATACCGCCCCTAGAACTAACATTC
fwd_TYR3_zebrafish	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACAACTGCTTTCCATGGTGTGT
rev_TYR3_zebrafish	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCCAGGGCTTTCGTGGAGA

Site	Sequence	SEQ ID NO
TYR1	$GTC_3AGGTC_8GAGGGTTCTGTCAGG$	564
TYR2	$CTTC_4C_5AGGATGAGAACACAGAGG$	565
TYR3	$CAAC_4C_5AC_7TGCTCAAAGATGCTGG$	566

Supplementary Table 1. Protospacer and PAM sequences for the zebrafish genomic loci studied in this work.

EXAMPLE 17: C:G-to-T:A Base Editors with Higher Efficiency and Product Purity

[00614] Base editing is the programmable conversion of target C:G base pairs to T:A base pairs without inducing double-stranded DNA breaks or requiring homology-directed repair using engineered fusions of Cas9 variants and cytidine deaminases (1). The third-generation base editor (BE3) and related technologies have been successfully used by many researchers in a wide range of organisms (2-13). The product distribution of base editing—the frequency with which

the target C:G base pair is converted to mixtures of undesired byproducts, along with the desired T:A product—varies in a target site-dependent manner (2, 3, 6-8). Here we characterize determinants of base editing outcomes in human cells, and establish that the formation of undesired products is dependent on uracil N-glycosylase (UNG), and is more likely to occur at target sites containing only a single C within the base editing activity window. The constructs CDA1-BE3 and AID-BE3, which use cytidine deaminase homologs that increase base editing efficiency for some sequences, were engineered. Additionally, a fourth-generation *S. pyogenes* Cas9-derived base editor (BE4) that more efficiently blocks access of UNG to base-edited intermediates was also engineered. Compared with BE3, BE4 increases by approximately 50% the efficiency of C:G to T:A base editing, while halving the frequency of undesired byproducts. These improvements were also applied to yield a *S. aureus* Cas9-derived BE4 (SaBE4), which is substantially smaller than BE4 and has an alternative targeting scope.

Introduction

[00615] Traditional genome editing methods introduce a double-stranded DNA break (DSB) at a genomic target locus (14). The cellular response to a DSB lesion primarily proceeds through nonhomologous end joining (NHEJ) and related processes (15). Although NHEJ usually rejoins the two ends flanking the DSB, under typical genome editing conditions DSBs are continuously reintroduced, eventually resulting in the accumulation of insertions and deletions (indels) or translocations at the site of the DSB and disruption of the corresponding genomic locus (16). Actively dividing cells can also respond to DSBs by initiating homology-directed repair (HDR) in the presence of a donor DNA template containing homology to the regions surrounding the DSB, which allows researchers to more precisely and predictably manipulate genomes than is possible through NHEJ (17). HDR-dependent genome editing is limited by low efficiency arising from competition with NHEJ outcomes, and from the dependence of HDR on mitosis (18).

[00616] The development of base editing, which enables the direct, irreversible conversion of a C:G base pair to a T:A base pair in a programmable manner without requiring HDR or the introduction of a DSB, has been reported (1). Base editors consist of a single-stranded DNA-specific cytidine deaminase enzyme tethered to a catalytically impaired Cas9 protein and a base excision repair inhibitor (1, 4, 9, 10). The Cas9 variant binds a genomic locus of interest, programmed by a corresponding guide RNA. Formation of the protein:RNA:DNA ternary "R-loop" complex (19) exposes a small (~5-nt) window of single-stranded DNA that serves as a

substrate for the tethered cytidine deaminase enzyme. Any cytidines within this window are hydrolytically deaminated to uracils, resulting in G:U intermediates.

[00617] Base excision repair (BER) is the cell's primary response to G:U mismatches and is initiated by excision of the uracil by uracil N-glycosylase (UNG)(20). In an effort to protect the edited G:U intermediate from excision by UNG, an 83-amino acid uracil glycosylase inhibitor (UGI) was fused directly to the C-terminus of catalytically dead Cas9 (dCas9) (1). To manipulate cellular DNA mismatch repair systems into preferentially replacing the G in the G:U mismatch with an A, the Ala 840 amino acid in dCas9 was reverted to His, enabling the Cas9 protein to nick the DNA strand opposite the newly formed uracil, resulting in much more efficient conversion of the G:U intermediate to desired A:U and A:T products (1). Combining these two engineering efforts resulted in BE3, a single protein consisting of a three-part fusion of the APOBEC1 cytidine deaminase enzyme tethered through a 16-amino acid linker to *S. pyogenes* dCas9(A840H), which is covalently linked to UGI through a 4-amino acid linker(l). BE3 and related base editors have now been employed for a wide variety of applications including plant genome editing, *in vivo* mammalian genome editing, targeted mutagenesis, and knockout studies (2-13). The scope of base editing has been recenelty expanded by reporting BE3 variants with altered PAM requirements (4), narrowed editing windows (4), reduced off-target editing (10), and small molecule dependence (21).

[00618] At some loci, base editors such as BE3 give rise to undesired byproducts in which the target C:G base pair is converted into a G:C or A:T base pair, rather than the desired T:A product (2, 3, 6-8). Here we illuminate determinants of base editing product purity, and establish that UNG activity is required for the formation of undesired byproducts. It has been determined that blocking UNG access to the uracil intermediate is especially crucial for target loci in which a single C is within the editing window in order to minimize undesired products. Fourth-generation base editors, BE4 and SaBE4, that perform base editing with higher efficiency and greatly improved product purity compared to previously described base editors including BE3 were engineered.

Results

UNG activity is requiredfor byproductformation

[00619] Undesired base editing byproducts may arise during base excision repair due to the formation and error-prone resolution of abasic sites within the uracil-containing DNA strand. To

determine if the product purity of base editing in cells lacking uracile N-glycosylase (UNG) improves, HAPl cells (a haploid human cell line) and HAPl UNG⁻ cells were nucelofected with plasmids encoding BE3 and sgRNAs targeting the *EMX1, FANCF, HEK2, HEK3, HEK4,* or *RNF2* loci (see **Figure 135B** for target sequences). Three days post-nucleofection, genomic DNA was extracted and the target loci were amplified by PCR and analyzed by high-throughput DNA sequencing (HTS). Base editing product purity is defined as the percent of edited sequencing reads (reads in which the target C has been converted to A, G, or T) in which the target C is edited to a T. The base editing product purity of BE3-treated HAPl cells averaged 68 ± 6% (mean ± S.D. for n=3 biological replicates) across 12 target Cs in the six loci. In HAPl UNG⁻ cells, all 12 target Cs tested were base edited with product purities > 98% **(Figure 135A).** In addition, indel frequencies at all six tested loci decreased 7- to 100-fold upon UNG knockout **(Figure 135C).** These data strongly implicate UNG activity as necessary for undesired product formation during base editing, consistent with a model in which abasic site formation and subsequent base excision repair with error-prone polymerases leads to randomization of the target nucleotide and occasional strand breaks that result in indels.

Targets with multiple editable Cs exhibit higher product purity

[00620] Base editing efficiency by BE3 can be lower for some (but not all) target Cs that are immediately downstream of a G (1), consistent with the known sequence preference of APOBEC1 (22) **(Figure 136A).** In an effort to efficiently edit such targets, BE3 variants in which replaced the APOBEC1 deaminase was replaced with CDA1 (to generate CDA1-BE3), AID (to generate AID-BE3), or APOBEC3G (to generate APOBEC3G-BE3), three single-stranded DNA-specific cytidine deaminase enzymes with different sequence preferences, were generated (23). HEK293T cells were transfected with plasmids encoding these BE3 variants and sgRNAs targeting the *EMX1, FANCF, HEK2, HEK3, HEK4,* or *RNF2* loci. Three days post-transfection, genomic DNA was extracted and the target loci were amplified by PCR and assessed for base editing using HTS. More efficient editing of target Cs that immediately follow a G was observed with CDA1-BE3 and AID-BE3 compared to BE3 **(Figure 136A, Figures 140A-D,** and **Figures 146-151).** In general, CDA1-BE3 and AID-BE3 exhibited lower editing efficiencies than BE3 at target Cs that do not follow a G **(Figures 140A-D).** In contrast, APOBEC3G-BE3 exhibited unpredictable sequence preferences, with overall lower yields of C-

to-T editing compared to BE3. These findings suggest that CDA1-BE3 and AID-BE3 may offer higher editing efficiencies over BE3 for some target 5'-GC-3' sequences.

[00621] While analyzing these data, it was noted that the product purities of CDA1-BE3 and AID-BE3 were typically higher than those of BE3 at those sites for which CDA1-BE3 and AID-BE3 edited more Cs than BE3 **(Figures 136A-D).** For example, at the *HEK4* locus, BE3 edits only a single C efficiently (the C not preceded by a G) but both CDA1-BE3 and AID-BE3 edit three Cs **(Figures 140A-C).** The product purity of BE3 at this locus is 50 ± 7% (mean ± S.D. for n = 3 biological replicates), while the product purity of CDA1-BE3 and AID-BE3 are 97 ± 2% and 93 ± 2%, respectively. Moreover, *EMXl* and *FANCF,* edited by BE3 with product purities of 84 ± 3% and 91 ± 2%, respectively, contain multiple Cs that are edited with comparable efficiency (Figure S2), while *HEK2* and *RNF2,* edited by BE3 with much lower product purities of 28 ± 3% and 64 ± 3%, respectively, contain multiple Cs that are edited with unequal efficiencies **(Figures 141A-C).** CDA1-BE3 and AID-BE3, which edit both Cs within the *HEK2* locus with comparable efficiencies, exhibit much higher product purities at this locus (85 ± 5% and 81 ± 4%, respectively) **(Figures 136A-D** and **Figure 140C).** The possibility that at the *HEK2* and *RNF2* sites the multiple Cs are initially converted to Us by BE3 with comparable efficiency and then processed with different efficiencies by DNA repair systems was ruled out. Given this, similar product distributions would be expected when these sites were treated with BE3 versus CDA1-BE3 or AID-BE3, rather than the different product distributions observed **(Figure 136B** and **Figures 146-151).** Instead, an isolated G:U may be more readily processed by UNG than clusters of G:U lesions. It is possible that the processivity of the cytidine deaminase domain in BE3 (1, 24) may increase the residence time of BE3 at loci containing multiple editable Cs, thereby blocking access by UNG more effectively than at loci containing a single editable C.

[00622] The relationship between product purity, the number of edited Cs in individual sequencing reads, and UNG activity was further analyzed. To reveal the fate of base edited DNA in the absence of explicit UNG inhibition, the UGI component of BE3 was removed to generate BE3B. HEK293T cells were transfected with plasmids encoding BE3 or BE3B and sgRNAs targeting the *EMXl, FANCF, HEK2, HEK3, HEK4,* or *RNF2* loci. As expected given the role of UNG in diversifying base editing outcomes established above, the product purities at all target

Cs greatly decreased in BE3B-treated DNA compared with BE3-treated DNA, with the fraction of editing products containing non-Ts increasing by an average of **1.8 ± 0**.4-fold **(Figure 142B).**

[00623] Individual DNA sequencing reads from HEK293T cells treated with sgRNAs targeting the multi-C sites *HEK2, HEK3,* and *RNF2* and either BE3 or BE3B were analyzed. For each site, the primary target C was designated as the nucleotide modified most efficiently. Across all three sites, an average of **80 ± 10%** of sequencing reads that contained an undesired C to non-T edit of the primary target C exhibited only that single base editing event **(Figures 142A-D** and **Figure 143).** In contrast, across the same three multi-C sites, a much lower average of 32 ± 4% of sequencing reads containing a clean C-to-T edit of the primary target C exhibited only that single clean base editing event **(Figures 142A-D** and **Figure 143).** In addition, the distribution of products for BE3B-treated HEK4 DNA, a site that contains only one C within the editing window, roughly follows the ratio of **1:3:1** for A:G:T **(Figure 143D).** These observations collectively indicate that when a single cytidine in a given target is converted to U in the absence of UGI, it is processed efficiently by UNG-initiated BER to give a mixture of products.

[00624] These data are consistent with a model in which clustered G:U mismatches are processed differently than isolated G:U mismatches, and are more likely to produce clean C-to-T edits. When only a single C-to-T editing event is desired, the above observations suggest that UNG inhibition is critical to minimize undesired byproducts. However, when performing targeted random mutagenesis using dCas9-deaminase fusions, such as with **TAM(8)** and CRISPR-X(2), the above observations suggest that target sites with only a single editable C will maximize product mixtures.

Optimization of BE3 architecture for improved product purity

[00625] The UGI component of BE3 was replaced with a single-stranded DNA binding (SSB) protein to yield SSB-BE3, such that SSB may block the uracil-containing ssDNA portion of the R-loop from being accessed by UNG. Large decreases in base editing efficiency by SSB-BE3 were observed, with all seven Cs across the four sites exhibiting an average of only **1.9 ± 0.5%** C-to-T conversion **(Figure 137C).**

[00626] Since the relative positioning of APOBEC, UGI, and UNG during steps that determine base editing outcomes are not known, UGI was relocated to the N-terminus of BE3 (N-UGI-BE3) in an effort to improve UNG inhibition. Moving UGI to the N-terminus of BE3 resulted in

an average decrease in C-to-T editing percentages across all seven tested target Cs of 2.3 ± 0.6-fold compared to BE3 **(Figure 137C),** and a decrease in overall product purity at all four sites compared to BE3 averaging 2.2 ± 0.5-fold **(Figure 137B).**

[00627] In contrast, appending an additional copy of UGI to the C-terminus of BE3 (BE3-2xUGI) resulted in large increases in product purities relative to BE3 and C-to-T editing percentages comparable to those of BE3. Non-T editing products decreased an average of 2.2 ± 0.8-fold across the four loci tested **(Figure 137B).** These observations suggest that addition of a second copy of UGI substantially decreases the access of UNG to the G:U base editing intermediate, thereby greatly improving product purity.

[00628] Because the above experiments also revealed the sensitivity of base editing outcomes to the architecture of the components, next we optimized the linkers between BE3 components to further increase product purities and editing efficiencies. We varied the rAPOBECl-dCas9(A840H) linker from 16 amino acids (BE3) to 32 amino acids (BE3C) and the dCas9(A840H)-UGI linker from 4 (BE3) to 9 (BE3D) to 16 amino acids (BE3E, **Figure 138A).** Non-T product formation on average decreased 1.3 ± 0.1-fold when the dCas9(A840H)-UGI linker was nine amino acid residues in length (BE3D) instead of four amino acids (BE3) **(Figure 138D),** with no apparent differences in C-to-T editing efficiencies **(Figure 138C).** Increasing the rAPOBECl-dCas9(A840H) linker from 16 amino acids (BE3) to 32 amino acids (BE3C) elevated C-to-T editing efficiencies an average of 1.2 ± 0.1-fold at the HEK2 locus **(Figure 138C).** This locus was previously the most unevenly edited multi-C site tested **(Figures 141A-C),** and extending this linker led to a reduction in preferential editing of C6 over C4 (the ratio of the percentage of sequencing reads that are edited at C6 to that of C4) from 2.6 ± 0.2-fold to 1.8 ± 0.1-fold. We reasoned that this longer linker may allow the deaminase better access to the ssDNA in the R-loop and result in more uniform deamination when multiple target Cs are present in the base editing window. BE3C also exhibited comparable or improved base editing efficiencies and product purities at the other loci tested **(Figures 138C-D).**

BE4, a C:G to T:A base editor with enhanced efficiency and product purity

[00629] The base editor construct BE4 was engineered by combining all three improvements—extending the rAPOBECl-dCas9 linker to 32 amino acids, extending the dCas9-UGI linker to 9 amino acids, and appending a second copy of UGI to the C-terminus of

the construct with another 9-amino acid linker. Target-AID, an alternative base editor construct reported by Nishida *et. al.* (9), was also cloned into the same plasmid backbone as BE4. HEK293T cells were transfected with plasmids encoding BE3, BE4, or Target-AID and sgRNAs targeting the *EMX1, FANCF, HEK2, HEK3, HEK4,* or *RNF2* loci. Three days post transfection, genomic DNA was extracted and the target loci were amplified by PCR and analyzed by HTS. An average increase in C-to-T editing efficiencies of 1.5 ± 0.3-fold across all twelve edited Cs for BE4 relative to BE3 was observed **(Figure 139C).** Although the average efficiency of C-to-T editing for Target-AID at the same positions analyzed was 1.5 ± 0.5-fold lower than that of BE3 and 2.1 ± 0.5-fold lower than that of BE4, it is important to note that Target-AID, which uses the CDA1 deaminase, appears to have an editing window shifted relative to BE3 and BE4, with optimal editing around positions C3 and C4 **(Figure 139C).** This shifted editing window makes comparisons of efficiency and product purity between Target-AID and BE3 or BE4 difficult because a given target C could lie in more optimal or less optimal position within the different editing windows, even when using the same guide RNA.

[00630] In addition to greater C-to-T editing efficiency, BE4 also exhibited substantially improved product purities relative to BE3 at all genomic loci tested, with an average decrease in non-T product formation of 2.3 ± 0.3-fold **(Figurel39D).** As expected from further impeding base excision repair, which can lead to indels (25), decreases in indel rates averaging 2.3 ± 1.1-fold across all six loci following BE4 treatment compared to BE3 were also observed **(Figures 144A-C).** Taken together, these results indicate that BE4 offers high efficiencies of C-to-T editing, high product purities, and low indel formation rates at all loci tested.

[00631] The BE4 improvements wereintegrated with *S. aureus* Cas9 (26) to generate SaBE4, which replaces the *S. pyogenes* dCas9(A840H) with the smaller *S. aureus* dCas9(A580N) and can access different targets due to its alternative PAM requirements. HEK293T cells were transfected with plasmids encoding SaBE3 (4) or SaBE4 and sgRNAs targeting the *FANCF, HEK3,* or *HEK4* loci. Consistent with the results comparing BE4 and BE3, we observed an average increase in C-to-T editing efficiencies of 1.4 ± 0.2-fold across all ten edited Cs for SaBE4 relative to SaBE3 **(Figure 145A),** with a 1.8 ± 0.5-fold average decrease in undesired non-T editing products **(Figure 145B).** These results indicate that the gains in base editing efficiency and product purity that arise from the BE4 enhancements also apply to base editors derived from other Cas9 homologs.

Materials and Methods

Cloning ofplasmids

[00632] All plasmids in this study were generated by USER cloning using Phusion U Hot Start polymerase (Thermo Fisher). Deaminase and SSB genes were synthesized as gBlocks Gene Fragments (Integrated DNA Technologies), and Target-AID was obtained from Addgene (plasmid # 79620). Protein sequences are listed in the Supplementary Notes.

Cell culture

[00633] HEK293T (ATCC CRL-3216) cells were maintained in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher) supplemented with 10% (v/v) fetal bovine serum (FBS), at 37 °C with 5% $C0_2$. HAP1 (Horizon Discovery C631) and HAP1 *UNG⁻* (Horizon Discovery HZGHC001531c012) were maintained in Iscove's Modified Dulbecco's Medium plus GlutaMax (ThermoFisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS), at 37 °C with 5% $C0_2$.

Transfections

[00634] HEK293T cells were seeded on 48-well collagen-coated BioCoat plates (Corning) and transfected at approximately 75% confluency. Briefly, 750 ng of BE and 250 ng of sgRNA expression plasmids were transfected using 1.5 μL of Lipofectamine 2000 (ThermoFisher Scientific) per well according to the manufacturer's protocol.

[00635] HAP1 and HAP1 *UNG⁻* cells were nucleofected using the SE Cell Line 4DNucleofector™ X Kit S (Lonza) according to the manufacturer's protocol. Briefly, 4×10^5 cells were nucleofected with 300 ng of BE and 100 ng of sgRNA expression plasmids using the 4DNucleofector™ program DZ-1 13.

High-throughput DNA sequencing of genomic DNA samples

[00636] Transfected cells were harvested after 3 days and the genomic DNA was isolated by incubating cells in lysis buffer (10 mM Tris-HCl pH 8.0, 0.05% SDS, 25 μg/mL proteinase K) at 37 °C for 1 hr followed by 80° C for 30 min. Genomic regions of interest were amplified by PCR with flanking HTS primer pairs as previously described (6, 1). PCR amplification was

carried out with Phusion high-fidelity DNA polymerase (ThermoFisher) according to the manufacturer's instructions and as previously described. Purified DNA was amplified by PCR with primers containing sequencing adaptors. The products were gel-purified and quantified using the QuantiT™ PicoGreen dsDNA Assay Kit (ThermoFisher) and KAPA Library Quantification Kit-Illumina (KAPA Biosystems). Samples were sequenced on an Illumina MiSeq as previously described.

Data analysis

[00637] Sequencing reads were automatically demultiplexed using MiSeq Reporter (Illumina), and individual FASTQ files were analyzed with a custom Matlab script as previously described (1). Each read was pairwise aligned to the appropriate reference sequence using the Smith-Waterman algorithm. Base calls with a Q-score below 31 were replaced with Ns and were thus excluded in calculating nucleotide frequencies. This treatment yields an expected MiSeq base-calling error rate of approximately 1 in 1,000. Aligned sequences in which the read and reference sequence contained no gaps were stored in an alignment table from which base frequencies could be tabulated for each locus.

[00638] Indel frequencies were quantified with the previously described Matlab script (5, 6, 1). Briefly, sequencing reads were scanned for exact matches to two 10-bp sequences that flank both sides of a window in which indels might occur. If no exact matches were located, the read was excluded from analysis. If the length of this indel window exactly matched the reference sequence the read was classified as not containing an indel. If the indel window was two or more bases longer or shorter than the reference sequence, then the sequencing read was classified as an insertion or deletion, respectively.

[00639] In order to evaluate interdependency (linkage disequilibrium) between the base editing outcomes at the multiple target cytidines within an editing window, target site sequences from BE treated cells were analyzed by a custom Python script (Supplementary Note 1). Briefly, sequencing reads were scanned for exact matches to two 7-bp sequences that flank each side of the protospacer. If the intervening region was not exactly 20-bp, then it was excluded further analysis. The protospacer sequences were further filtered into four groups based upon the identity of the nucleotide at the position with the most non-T editing outcomes (the primary

target C). For each of these four groups as well as the entire pool, the nucleotide abundance at each of the 20 positions within the protospacer were tallied.

EXAMPLE 18: Base Editors Comprising an LbCpfl (nuclease dead, nuclease active, and nickase).

[00640] As discussed above, nucleic acid programmable DNA binding proteins (napDNAbp) of any of the fusion proteins provided herein may be an LbCpfl protein. In some embodiments, the LbCpfl protein is nuclease inactive, nuclease active, or an LbCpfl nickase. Several constructs of fusion proteins comprising forms of LbCpfl were tested for their ability to make C to T edits in different target sequences. A schematic representation of the constructs tested is shown in Figure 152. Construct 10 has a domain arrangement of [Apobec]-[LbCpfl]-[UGI]-[UGI]; construct 11 has a domain arrangement of [Apobec]-[LbCpfl]-[UGI]; construct 12 has a domain arrangement of [UGI]-[Apobec]-[LbCpfl]; construct 13 has a domain arrangement of [Apobec]-[UGI]-[LbCpfl]; construct 14 has a domain arrangement of [LbCpfl]-[UGI]-[Apobec]; construct 15 has a domain arrangement of [LbCpfl]-[Apobec]-[UGI]. For each construct three different LbCpfl proteins were used (D/N/A, which refers to nuclease dead LbCpfl (D); LbCpfl nickase (N) and nuclease active LbCpfl (A)). For each of these constructs, the linkers linking the domains are shown below, where XTEN refers to the XTEN linker having the sequence SGSETPGTSESATPES (SEQ ID NO: 604), and BPNLS refers to the nuclear localization sequence having the sequence KRTADGSEFEPKKKRKV (SEQ ID NO: 740). Constructs are shown from N-terminus (left) to C-terminus (right).

Construct 10: **Apobec-SGGSSGGSXr *EN*SGGSSGGS-LbCpfl-SGGSGGSGGS-UGI-SGGSGGSGGS-UGI-SGGS-BPNLS**

Construct 11: **Apobec-SGGSSGGSXr *EN*SGGSSGGS-LbCpfl-SGGSGGSGGS-UGI-SGGS**-BPNLS

Construct 12: UGI-SGGSGGSGGS-Apobec-SGGSSGGSXTWSGGSSGGS-LbCpfl-SGGS-BPNLS

Construct 13: Apobec-SGGSGGSGGS-UGI-SGGSSGGSXTWSGGSSGGS-LbCpfl-SGGS-BPNLS

Construct 14: LbCpfl-SGGSGGSGGS-UGI-SGGSGGSGGS-Apobec-SGGS-BPNLS

Construct 15: LbCpfl-SGGSGGSGGS-Apobec-SGGSGGSGGS-UGI-SGGS-BPNLS

[00641] **The** common guide backbone (sgRNA) used in the experiments is GTAATTTCTACTAAGTGTAGAT (SEQ ID NO: 741)[guide sequence]TTTTTTT, wherein each of the Ts of SEQ ID NO: 741 are uracil (U), and where the guide sequence that targets the construct to a specific nucleotide sequence is shown between brackts. In some embodiments, any of the constructs provided herein are complexed with a sgRNA that comprises the backbone sequence of GTAATTTCTACTAAGTGTAGAT(SEQ ID NO: 741), wherein each of the Ts of SEQ ID NO: 741 are uracil (U). In some embodiments, any of the guide RNAs provided herein comprise a guide sequence comprising 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides that are perfectly complementary to a sequence, *e.g.,* a target DNA sequence. In the experiments performed, the guide sequences tested are shown below:

EMX23: TACTTTGTCCTCCGGTTCTGGAA (SEQ ID NO: 742)

EMX20: TACTTTGTCCTCCGGTTCTG (SEQ ID NO: 743)

EMX19: TACTTTGTCCTCCGGTTCT (SEQ ID NO: 744)

EMX18: TACTTTGTCCTCCGGTTC (SEQ ID NO: 745)

EMX17: TACTTTGTCCTCCGGTT (SEQ ID NO: 746)

Hek2_23: CAGCCCGCTGGCCCTGTAAAGGA (SEQ ID NO: 747)

Hek2_20: CAGCCCGCTGGCCCTGTAAA (SEQ ID NO: 748)

Hek2_19: CAGCCCGCTGGCCCTGTAA (SEQ ID NO: 749)

Hek2_18: CAGCCCGCTGGCCCTGTA (SEQ ID NO: 750)

Hek2_17: CAGCCCGCTGGCCCTGT (SEQ ID NO: 751)

[00642] The data demonstrating the C to T base pair editing percentage using various constructs and target sequences is shown in figures 153-159, and the editing percentage values (after adjustment based on indel count), and the percentage of indels for the experiments are shown in figures 160-166.

Supplementary Sequences

[00643] Amino Acid Sequences of CDA1-BE3, AID-BE3, BE4, and SaBE4 fusion proteins.

CDA1-XTEN-dCas9-UGI-*NLS* primary sequence (CDA1-BE3):

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTE
RGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACK
LYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKR
*RSELSIMIQVKILHTTKSPAV*SGSETPGTSESATPES DKKYSIGLAIGTNSVGWAVITDEYK

VPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF
SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGV
DAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGT
EELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDS LTFKEDIQKAQ VSGQGDSLHEHIANLAGS PAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN
TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSD
KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERS SFEKNPIDFLE AKGYKEVKKDLIIKLPKYS LFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKOLVIOESILMLP
EEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIK
MLSGGSPKKKRKV (SEQ ID NO: 165)

A*HD*-XTEN-**dCas9**-UGI-*NLS* **primary sequence** (AID-BE3):

MDSLLMNRRKFLYOFKNVRWAKGRRETYLCYWKRRDSA TSFSLDFGYLRNKNGCHVELLFL
RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPE
GLRRLHRAGVOIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSROLRRILLPLYEVD
DL^DAF^rLGLSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKV
LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSA
RLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL
LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY
EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY

LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY
LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL
ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGN**
KPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSP*KKK*
RKV (SEQ ID NO: 166)

rAPOBECl **-linker-dCas9-UGI-***UGI-NLS* primary sequence (BE4):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVE
VNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNR
QGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPC
LNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL
AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH
QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGDSGGSGGSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA**
YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGS77VL5D *HEK*
ETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWAL
VIQDSNGENKIKMLS*GGSPKKKRK* (SEQ ID NO: 167)

rAPOBECl **-linker-SaCas9d-UGI-*UGI*-NLS** primary sequence (SaBE4):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSONTNKHVE VNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNR OGLRDLISSGVTIOIMTEOESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPC LNILRRKOPOLTFFTIALOSCHYORLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESS GGSSGGSGKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQR VKKLLFD YNLLTDHS ELS GINPYEARVKGLS QKLS EEEFS AALLHLA KRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFK TSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYE MLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQ KKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIA KILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQ IAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFN YEVDHIIPRS VSFDNS FNNKVLVKQEENS KKGNRTPFQ YLS SSDSK ISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMN LLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKE WKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVD KKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDP QTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLD ITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEA KKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMND KRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGGSPKKKRKVSSDYKDHDG DYKDHDID YKDDDDKS GGS GGS GGS **TNLSDIIEKETGKQLVIQESILMLPEE VEEVIG NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGG** SGGS ***TNLSDIIEKETGKQL VIQESILMLPEE VEE VIGNKPESDIL VHTA YDESTDENVMLL TSDAPE YKPWALVIQDSNGENKIKMLS*** *GGSPKKKRKV* (SEQ ID NO: 168)

References for Example 17

1. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420-424 (2016).

2. G. T. Hess et al., Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat Methods 13, 1036-1042 (2016).

3. K. Kim et al., Highly efficient RNA-guided base editing in mouse embryos. Nat Biotechnol 35, 435-437 (2017).

4. Y. B. Kim et al., Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat Biotechnol 35, 371-376 (2017).

5. C. Kuscu et al., CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. Nat Meth advance online publication, (2017).

6. J. Li, Y. Sun, J. Du, Y. Zhao, L. Xia, Generation of Targeted Point Mutations in Rice by a Modified CRISPR/Cas9 System. Mol Plant 10, 526-529 (2017).

7. Y. Lu, J. K. Zhu, Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. Mol Plant 10, 523-525 (2017).

8. Y. Ma et al., Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat Methods 13, 1029-1035 (2016).

9. K. Nishida et al., Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353, (2016).

10. H. A. Rees et al., Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun 8, 15790 (2017).

11. L. Yang et al., Engineering and optimising deaminase fusions for genome editing. Nat Commun 7, 13330 (2016).

12. Y. Zong et al., Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat Biotechnol 35, 438-440 (2017).

13. Z. Shimatani et al., Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat Biotech 35, 441-443 (2017).

14. A. C. Komor, A. H. Badran, D. R. Liu, CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell 168, 20-36 (2017).

15. A. J. Davis, D. J. Chen, DNA double strand break repair via non-homologous end-joining. Translational cancer research 2, 130-143 (2013).

16. M. M. Vilenchik, A. G. Knudson, Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. Proceedings of the National Academy of Sciences 100, 12871-12876 (2003).

17. F. Liang, M. Han, P. J. Romanienko, M. Jasin, Homology-directed repair is a major doublestrand break repair pathway in mammalian cells. Proceedings of the National Academy of Sciences 95, 5172-5177 (1998).

18. Y. Miyaoka et al., Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. Scientific Reports 6, 23549 (2016).

19. M. M. Jore et al., Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol 18, 529-536 (2011).

20. L. H. Pearl, Structure and function in the uracil-DNA glycosylase superfamily. Mutation Research/DNA Repair 460, 165-181 (2000).

21. W. Tang, J. H. Hu, D. R. Liu, Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. 8, 15939 (2017).

22. G. Saraconi, F. Severi, C. Sala, G. Mattiuz, S. G. Conticello, The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. Genome Biology 15, 417 (2014).

23. R. M. Kohli et al., Local Sequence Targeting in the AID/APOBEC Family Differentially Impacts Retroviral Restriction and Antibody Diversification. Journal of Biological Chemistry 285, 40956-40964 (2010).

24. L. Chelico, P. Pham, M. F. Goodman, Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G. Philosophical Transactions of the Royal Society B: Biological Sciences 364, 583-593 (2009).

25. E. A. Kouzminova, A. Kuzminov, Patterns of chromosomal fragmentation due to uracil-DNA incorporation reveal a novel mechanism of replication-dependent double-stranded breaks. Molecular Microbiology 68, 202-215 (2008).

26. F. A. Ran et al., In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186-191 (2015).

27. F. d. A. di Fagagna, G. R. Weller, A. J. Doherty, S. P. Jackson, The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku. EMBO Reports 4, 47-52 (2003).

28. C. Shee et al., Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. eLife 2, e01222 (2013).

EQUIVALENTS AND SCOPE

[00644] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents of the embodiments described herein. The scope of the present disclosure is not intended to be limited to the above description, but rather is as set forth in the appended claims.

[00645] Articles such as "a," "an," and "the" may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include "or" between two or more members of a group are considered satisfied if one, more than one, or all of the group members are present, unless indicated to the contrary or otherwise evident from the context. The disclosure of a group that includes "or" between two or more group members provides embodiments in which exactly one member of the group is present, embodiments in which more than one members of the group are present, and embodiments in which all of the group members are present. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[00646] It is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitation, element, clause, or descriptive term, from one or more of the claims or from one or more relevant portion of the description, is introduced into another claim. For example, a claim that is dependent on another claim can be modified to include one or more of the limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of making or using the composition according to any of the methods of making or using disclosed herein or according to methods known in the art, if any, are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[00647] Where elements are presented as lists, *e.g.,* in Markush group format, it is to be understood that every possible subgroup of the elements is also disclosed, and that any element or subgroup of elements can be removed from the group. It is also noted that the term "comprising" is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where an embodiment, product, or method is referred to as comprising particular elements, features, or steps, embodiments, products, or methods that consist, or consist essentially of, such elements, features, or steps, are provided as well. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[00648] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in some embodiments, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. For purposes of brevity, the values in each range have not been individually spelled out herein, but it will be understood that each of these values is provided herein and may be specifically claimed or disclaimed. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.
In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims. For purposes of brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.

CLAIMS

What is claimed is:

1. A fusion protein comprising: (i) a nucleic acid programmable DNA binding protein (napDNAbp); (ii) a cytidine deaminase domain; and (iii) a uracil glycosylase inhibitor (UGI) domain, wherein the napDNAbp is a CasX, CasY, Cpfl, C2cl, C2c2, C2c3, or Argonaute protein.

2. The fusion protein of claim 1, wherein the fusion protein comprises two, three, four, or five UGI domains

3. The fusion protein of claim 1 or 2, wherein the nucleic acid programmable DNA binding protein (napDNAbp) is a CasX protein.

4. The fusion protein of claim 3, wherein the CasX protein comprises
(i) an amino acid sequence that is at least 90% identical to SEQ ID NO: 29 or 30; or
(ii) the amino acid sequence of SEQ ID NO: 29 or 30.

5. The fusion protein of claim 1 or 2, wherein the nucleic acid programmable DNA binding protein (napDNAbp) is a CasY protein.

6. The fusion protein of claim 5, wherein the CasY protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 31.

7. The fusion protein of claim 5 or 6, wherin the CasY protein comprises the amino acid sequence of SEQ ID NO: 31.

8. The fusion protein of claim 1 or 2, wherein the nucleic acid programmable DNA binding protein (napDNAbp) is a Cpfl or Cpfl mutant protein.

9. The fusion protein of claim 8, wherein the Cpf1 or Cpf1 mutant protein comprises an amino acid sequence that is at least 90% identical to any one of SEQ ID NOs: 9-16.

10. The fusion protein of claim 8 or 9, wherin the Cpf1 or Cpfl mutant protein comprises the amino acid sequence of any one of SEQ ID NOs: 9-16.

11. The fusion protein of claim 1 or 2, wherein the nucleic acid programmable DNA binding protein (napDNAbp) is a C2cl protein.

12. The fusion protein of claim 11, wherein the C2cl protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 26.

13. The fusion protein of claim 11 or 12, wherin the C2cl protein comprises the amino acid sequence of SEQ ID NO: 26.

14. The fusion protein of claim 1 or 2, wherein the nucleic acid programmable DNA binding protein (napDNAbp) is a C2c2 protein.

15. The fusion protein of claim 14, wherein the C2c2 protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 27.

16. The fusion protein of claim 14 or 15, wherin the C2c2 protein comprises the amino acid sequence of SEQ ID NO: 27.

17. The fusion protein of claim 1 or 2, wherein the nucleic acid programmable DNA binding protein (napDNAbp) is a C2c3 protein.

18. The fusion protein of claim 17, wherein the C2c3 protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 28.

19. The fusion protein of claim 17 or 18, wherin the C2c3 protein comprises the amino acid sequence of SEQ ID NO: 28.

20. The fusion protein of claim 1 or 2, wherein the nucleic acid programmable DNA binding protein (napDNAbp) is an Argonaute protein.

21. The fusion protein of claim 20, wherein the Argonaute protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 25.

22. The fusion protein of claim 20 or 21, wherin the Argonaute protein comprises the amino acid sequence of SEQ ID NO: 25.

23. The fusion protein of any one of claims 1-22, wherein the cytidine deaminase domain is a deaminase from the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase.

24. The fusion protein of claim 23, wherein the APOBEC family deaminase is selected from the group consisting of APOBECl deaminase, APOBEC2 deaminase, APOBEC3A deaminase, APOBEC 3B deaminase, APOBEC 3C deaminase, APOBEC 3D deaminase, APOBEC 3F deaminase, APOBEC3G deaminase, and APOBEC3H deaminase.

25. The fusion protein of any one of claims 1-24, wherein the cytidine deaminase domain comprises an amino acid sequence that is at least 85% identical to an amino acid sequence of SEQ ID NO: 49-84.

26. The fusion protein of any one of claims 1-25, wherein the cytidine deaminase domain comprises an amino acid sequence of SEQ ID NO: 49-84.

27. The fusion protein of any one of claims 1-22, wherein the cytidine deaminase domain is a rat APOBECl (rAPOBECl) deaminase comprising one or more mutations selected from the group consisting of W90Y, R126E, and R132E of SEQ ID NO: 76, or one or more corresponding mutations in another APOBEC deaminase.

28. The fusion protein of any one of claims 1-22, wherein the cytidine deaminase domain is a human APOBECl (hAPOBECl) deaminase comprising one or more mutations selected from the group consisting of W90Y, Q126E, and R132E of SEQ ID NO: 74, or one or more corresponding mutations in another APOBEC deaminase.

29. The fusion protein of any one of claims 1-22, wherein the cytidine deaminase domain is a human APOBEC3G (hAPOBEC3G) deaminase comprising one or more mutations selected from the group consisting of W285Y, R320E, and R326E of SEQ ID NO: 60, or one or more corresponding mutations in another APOBEC deaminase.

30. The fusion protein of any one of claims 1-22, wherein the cytidine deaminase domain is an activation-induced deaminase (AID).

31. The fusion protein of any one of claims 1-22, wherein the cytidine deaminase domain is a cytidine deaminase 1 from *Petromyzon marinus* (pmCDAl).

32. The fusion protein of any one of claims 1-31, wherein the UGI domain comprises a domain capable if inhibiting UDG activity.

32. The fusion protein of any one of claims 1-32, wherein the UGI domain comprises an amino acid sequence that is at least 85% identical to SEQ ID NO: 134.

33. The fusion protein of any one of claims 132, wherein the UGI domain comprises an amino acid sequence as set forth in SEQ ID NO: 134.

34. The fusion protein of any one of claims 1-33, wherein the fusion protein comprises the structure:

NH_2-[cytidine deaminase domain] -[napDNAbp]- [UGI domain]-COOH;
NH_2-[cytidine deaminase domain]-[napDNAbp]-[UGI]-[UGI]-COOH;
NH_2-[cytidine deaminase domain]-[napDNAbp]-[UGI]-COOH;

NH_2-[UGI]-[Apobec]-[napDNAbp]-COOH;

NH_2-[cytidine deaminase domain]-[UGI]-[napDNAbp]-COOH;

NH_2-[napDNAbp]-[UGI]-[cytidine deaminase domain]-COOH; or

NH_2-[napDNAbp]-[cytidine deaminase domain]-[UGI] -COOH;

wherein each instance of "-" comprises an optional linker.

35. The fusion protein of any one of claims 34, wherein the cytidine deaminase domain of (ii) and the napDNAbp domain of (i) are linked via a linker comprising the amino acid sequence $(GGGS)_n$ (SEQ ID NO: 613), $(GGGGS)_n$ (SEQ ID NO: 607), $(G)_n$ (SEQ ID NO: 608), $(EAAAK)_n$ (SEQ ID NO: 609), $(GGS)_n$ (SEQ ID NO: 610), $(SGGS)_n$ (SEQ ID NO: 606), SGSETPGTSESATPES (SEQ ID NO: 604), or $(XP)_n$ (SEQ ID NO: 611) motif, or a combination thereof, wherein $_n$ is independently an integer between 1 and 30, inclusive, and wherein X is any amino acid.

36. The fusion protein of any one of claims 1-35, wherein the cytidine deaminase domain of (ii) and the napDNAbp of (i) are linked via a linker comprising the amino acid sequence: SGSETPGTSESATPES (SEQ ID NO: 604).

37. The fusion protein of any one of claims 1-36 further comprising a nuclear localization sequence (NLS).

38. The fusion protein of claim 37, wherein the NLS comprises the amino acid sequence PKKKRKV (SEQ ID NO: 614), MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 615), or KRTADGSEFEPKKKRKV (SEQ ID NO: 740)

39. The fusion protein of claim 37 or 38, wherein the fusion protein comprises the structure: NH_2-[cytidine deaminase domain]-[napDNAbp]-[UGI domain]-[NLS]-COOH, and wherein each instance of "-" comprises an optional linker.

40. The fusion protein of any one of claims 37-39, wherein the UGI domain and the NLS are linked via a linker comprising the amino acid sequence: SGGS (SEQ ID NO: 606), or wherein

the napDNAbp and the UGI domain are linked via a linker comprising the amino acid sequence: SGGS (SEQ ID NO: 606).

41 A complex comprising the fusion protein of any one of claims 1-40 and a guide RNA bound to the napDNAbp of the fusion protein.

42. A method comprising contacting a nucleic acid molecule with the fusion protein of any one of claims 1-40 and a guide RNA, wherein the guide RNA comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence in the genome of an organism and comprises a target base pair.

43. The method of claim 42, wherein the target base pair comprises a T to C point mutation associated with a disease or disorder, and wherein the deamination of the mutant C base results in a sequence that is not associated with a disease or disorder.

44. The method of claim 42 or 43, wherein the contacting results in less than 20% indel formation upon base editing.

45. The method of any one of claims 42-44, wherein the contacting results in at least 2:1 intended to unintended product upon base editing.

46. A complex comprising the fusion protein of anyone of claims 1-40, and an RNA bound to the napDNAbp.

47. The complex of claim 46, wherein the RNA is a guide RNA (gRNA).

48. The complex of claim 46, wherein the RNA is a non-coding RNA, and/or wherein the RNA comprises the amino acid sequence of SEQ ID NO: 741, wherein each of the Ts of SEQ ID NO: 741 are uracil (U).

49. The complex of claim 46, wherein the RNA is a microRNA (miRNA).

50. The complex of claim 46, wherein the RNA is a small interfering RNA (siRNA).

51. The complex of claim 46, wherein the RNA is a Piwi-interacting RNA (piRNA).

52. The complex of any one of claims 46-51, wherein the RNA is from 10-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence.

53. The complex of claim 52, wherein the RNA is 10, 11, 12, 13, 14,15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides long.

54. The complex of any one of claims 46-53, wherein the RNA comprises a sequence of 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 contiguous nucleotides that is complementary to a target sequence.

55. The complex of any one of claims 46-54, wherein the target sequence is a DNA sequence.

56. The complex of any one of claims 46-54, wherein the target sequence is a RNA sequence.

57. The complex of claim 55, wherein the target sequence is in the genome of an organism.

58. The complex of claim 56, wherein the target sequence is transcribed from the genome of an organism.

59. The complex of claim 57 or 58, wherein the organism is a prokaryote.

60. The complex of claim 59, wherein the prokaryote is bacteria.

61. The complex of claim 57 or 58, wherein the organism is a eukaryote.

62. The complex of claim 61, wherein the organism is a plant.

63. The complex of claim 61, wherein the organism is a vertebrate.

64. The complex of claim 63, wherein the vertebrate is a mammal.

65. The complex of claim 64, wherein the mammal is a mouse or rat.

66. The complex of claim 64, wherein the mammal is human.

67. A method comprising contacting a nucleic acid molecule with the complex of any one of claims 46-66.

68. The method of claim 67, wherein the nucleic acid is DNA.

69. The method of claim 68, wherein the nucleic acid is double-stranded DNA.

70. The method of claim 67, wherein the nucleic acid is RNA.

71. The method of claim 70, wherein the RNA is mRNA.

72. The method of any one of claims 67-71, wherein the nucleic acid comprises a target sequence associated with a disease or disorder.

73. The method of claim 72, wherein the target sequence comprises a point mutation associated with a disease or disorder.

74. The method of claim 73, wherein the activity of the fusion protein, or the complex results in a correction of the point mutation.

75. The method of any one of claims 73-74, wherein the target sequence comprises a T to C point mutation associated with a disease or disorder, and wherein the deamination of the mutant C base results in a sequence that is not associated with a disease or disorder.

76. The method of claim 75, wherein the target sequence encodes a protein, and wherein the point mutation is in a codon and results in a change in the amino acid encoded by the mutant codon as compared to a wild-type codon.

77. The method of claim 75 or 76, wherein the deamination of the mutant C results in a change of the amino acid encoded by the mutant codon.

78. The method of claim 75 or 76, wherein the deamination of the mutant C results in the codon encoding a wild-type amino acid.

79. The method of any one of claims 67-78, wherein the contacting is performed *in vivo* in a subject.

80. The method of any one of claims 67-78, wherein the contacting is performed *in vitro.*

81. The method of claim 79, wherein the subject has been diagnosed with a disease or disorder.

82. The method of any one of claims 72-82, wherein the disease or disorder is cystic fibrosis, phenylketonuria, epidermolytic hyperkeratosis (EHK), Charcot-Marie-Toot disease type 4J, neuroblastoma (NB), von Willebrand disease (vWD), myotonia congenital, hereditary renal amyloidosis, dilated cardiomyopathy (DCM), hereditary lymphedema, familial Alzheimer's disease, HIV, Prion disease, chronic infantile neurologic cutaneous articular syndrome (CINCA),

desmin-related myopathy (DRM), a neoplastic disease associated with a mutant PI3KCA protein, a mutant CTNNB 1 protein, a mutant HRAS protein, or a mutant p53 protein.

83. A kit comprising a nucleic acid construct, comprising
(a) a nucleic acid sequence encoding the fusion protein of any one of claims 1-40; and
(b) a heterologous promoter that drives expression of the sequence of (a).

84. The kit of claim 83, further comprising an expression construct encoding a guide RNA backbone, wherein the construct comprises a cloning site positioned to allow the cloning of a nucleic acid sequence identical or complementary to a target sequence into the guide RNA backbone.

85. A polynucleotide encoding the fusion protein of any one of claims 1-40.

86. A vector comprising a polynucleotide of claim 85.

87. The vector of claim 86, wherein the vector comprises a heterologous promoter driving expression of the polynucleotide.

88. A cell comprising the fusion protein of any one of claims 1-40.

89. A cell comprising the complex of any of claims 46-66.

90. A cell comprising the nucleic acid molecule encoding the fusion protein of any one of claims 1-40.

91. A method for producing a ribonucleoprotein (RNP) complex, the method comprising: (i) complexing a base editor protein with an RNA in an aqueous solution, thereby forming a complex comprising the base editor and the RNA in the aqueous solution; and
(ii) contacting the complex of (i) with a cationic lipid.

92. The method of claim 91, wherein the base editor protein and the RNA of (i) are complexed at a molar ratio from 1:1 to 1:1.5.

93. The method of claim 91 or 92, wherein the base editor protein and the RNA of (i) are complexed at a molar ratio from 1:1.05 to 1:1.2.

94. The method of any one of claims 91-93, wherein the base editor protein and the RNA of (i) are complexed at a molar ratio of about 1:1.1.

95. The method of any one of claims 91-94, wherein the base editor protein and the RNA of (i) are complexed at a molar ratio of 1:1.1.

96. The method of any one of claims 91-95, wherein the base editor protein is in the aqueous solution at a concentration from 10 μM and 100 μM.

97. The method of any one of claims 91-96, wherein the base editor protein is in the aqueous solution at a concentration from 20 μM and 50 μM.

98. The method of any one of claims 91-97, wherein the base editor protein is in the aqueous solution at a concentration from 30 μM and 40 μM.

99. The method of any one of claims 91-98, wherein the base editor protein is in the aqueous solution at a concentration from 11 μM and 110 μM.

100. The method of any one of claims 91-99, wherein the base editor protein is in the aqueous solution at a concentration from 22 μM and 55 μM.

101. The method of any one of claims 91-100, wherein the base editor protein is in the aqueous solution at a concentration from 33 μM and 44 μM.

102. The method of any one of claims 91-101, wherein the RNA is a sgRNA.

103. The method of any one of claims 91-102, wherein the complex in the aqueous solution of (i) is contacted with the cationic lipid of (ii) at a volumetric ratio that is from 1:2 to 2:1.

104. The method of any one of claims 91-103, wherein the complex in the aqueous solution of (i) is contacted with the cationic lipid of (ii) at a volumetric ratio that is from 1:1.5 to 1.5:1.

105. The method of any one of claims 91-104, wherein the complex in the aqueous solution of (i) is contacted with the cationic lipid of (ii) at a volumetric ratio that is from 1:1.2 to 1.2:1.

106. The method of any one of claims 91-105, wherein the complex in the aqueous solution of (i) is contacted with the cationic lipid of (ii) at a volumetric ratio that is from 1:1.1 to 1.1:1.

107. The method of any one of claims 91-106, wherein the complex in the aqueous solution of (i) is contacted with the cationic lipid of (ii) at a volumetric ratio that is about 1:1.

108. The method of any one of claims 91-107, wherein the complex in the aqueous solution of (i) is contacted with the cationic lipid of (ii) at a volumetric ratio that is 1:1.

109. The method of any one of claims 91-108, wherein the cationic lipid is Lipofectamine®.

110. The method of claim 109, wherein the Lipofectamine® is selected from the group consisting of Lipofectamine® 2000, Lipofectamine®3000, Lipofectamine® MessengerMAX, Lipofectamine® LTX, and Lipofectamine®RNAiMAX.

111. The method of any one of claims 91-1 10, wherein the base editor is the fusion protein of any one of claims 1-40.

112. The method of any one of claim 91-111, wherein the complex of (i) is the complex of any one of claims 46-66.

113. A pharmaceutical composition produced by the method of any one of claims 91-112.

114. A pharmaceutical composition comprising the fusion protein of any one of claims 1-40.

115. A pharmaceutical composition comprising the complex of claim 41.

116. A pharmaceutical composition comprising the complex of any one of claims 46-66.

117. The pharmaceutical composition of any one of claims 113-116 further comprising a pharmaceutically acceptable excipient.

118. A method for purifying a base editor protein, the method comprising:
(i) expressing the base editor protein in a cell, wherein the base editor protein comprises an affinity tag;
(ii) lysing the cell of (i), thereby generating a lysate; and
(iii) subjecting the lysate to affinity chromatography, thereby producing an eluent comprising a purified base editor protein.

119. The method of claim 118, wherein the cell is a bacterial cell.

120. The method of claim 119, wherein the bacterial cell is an *E. coli* cell.

121. The method of claim 120, wherein the *E. coli* cell is a BL21 DE3 competent cell.

122. The method of any one of claims 118-121, wherein the affinity tag is a polyhistidine tag.

123. The method of any one of claims 118-122, wherein step (iii) comprises contacting the lysate of (ii) with a nickel-NTA resin, wherein the base editor protein is bound to the nickel-NTA resin.

124. The method of claim 123, wherein the method further comprises

(iv) eluting the base editor protein from the nickel-NTA resin, thereby generating an eluent comprising the base editor protein.

125. The method of any one of claims 118-124, wherein the eluent comprising the base editor protein is subjected to cation exchange chromatography.

126. The method of claim 125, wherein the cation exchange chromatography comprises contacting the eluent comprising the base editor protein with a cation exchange column.

127. A method comprising delivering the fusion protein of any one of claims 1-40, the complex of claim 41, the complex of any one of claims 46-66, or the pharmaceutical composition of any one of claims 113-1 17 to the inner ear of a subject.

128. The method of claim 127 comprising delivering the fusion protein of any one of claims 1-40, the complex of claim 41, the complex of any one of claims 46-66, or the pharmaceutical composition of any one of claims 113-1 17 to the cochlea of the subject.

129. The method of claim 127 or 128, wherein the fusion protein of any one of claims 1-40, the complex of claim 41, the complex of any one of claims 46-66, or the pharmaceutical composition of any one of claims 113-1 17 is injected into the inner ear of the subject.

130. The method of any one of claims 127-129, wherein the fusion protein of any one of claims 1-40, the complex of claim 41, the complex of any one of claims 46-66, or the pharmaceutical composition of any one of claims 113-117 is injected into the cochlea of the subject.

131. The method of any one of claims 127-130, further comprising making a postauricular incision into the ear of the subject.

132. A method comprising delivering the fusion protein of any one of claims 1-40, the complex of claim 41, the complex of any one of claims 46-66, or the pharmaceutical composition of any one of claims 113-1 17 to a zebrafish embryo.

133. The method of claim 132, wherein the fusion protein of any one of claims 1-40, the complex of claim 41, the complex of any one of claims 46-66, or the pharmaceutical composition of any one of claims 113-1 17 is injected into the zebrafish embryo at the one-cell stage.

134. The method of claim 132 or 133, wherein the fusion protein of any one of claims 1-40, the complex of claim 41, the complex of any one of claims 46-66, or the pharmaceutical composition of any one of claims 113-1 17 is injected into the zebrafish embryo at the two-cello, four-cell, eight-cell, or sixteen-cell stage.

135. The method of any one of claims 132-134, wherein between 0.5 nmol and 50 nmol of base editor protein complexed with a gRNA is injected into the zebrafish embryo.

136. The method of any one of claims 132-135, wherein between 2 nmol and 30 nmol of base editor protein complexed with a gRNA is injected into the zebrafish embryo.

137. A fusion protein comprising: (i) a nucleic acid programmable DNA binding protein (napDNAbp); (ii) a cytidine deaminase domain; (iii) a first uracil glycosylase inhibitor (UGI) domain; and (iv) a second uracil glycosylase inhibitor (UGI) domain.

138. The fusion protein of claim 137, wherein the nucleic acid programmable DNA binding protein (napDNAbp) is is a CasX, CasY, Cpfl, Cpfl nickase, dCpfl, C2cl, C2c2, C2c3, Cas9, dCas9, Cas9 nickase or Argonaute protein.

139. The fusion protein of claim 138, wherein the dCas9 protein comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 6 or 7.

140. The fusion protein of claim 137 or 138, wherein the dCas9 protein comprises the amino acid sequence of SEQ ID NO: 7.

141. The fusion protein of claim 138, wherein the dCas9 protein is a nuclease inactive *Staphylococcus aureus* Cas9 (SaCas9d) protein.

142. The fusion protein of claim 141, wherein the SaCas9d protein comprises an amino acid sequence that is at least 85%, 90%, 95%, or 98% identical to SEQ ID NO: 33-36.

143. The fusion protein of claim 141 or 142, wherein the SaCas9d protein comprises the amino acid sequence of SEQ ID NO: 34.

144. The fusion protein of any one of claims 137-143, wherein the cytidine deaminase domain is a deaminase from the apolipoprotein B mRNA-editing complex (APOBEC) family deaminase.

145. The fusion protein of claim 144, wherein the APOBEC family deaminase is selected from the group consisting of APOBECl deaminase, APOBEC2 deaminase, APOBEC3A deaminase, APOBEC 3B deaminase, APOBEC 3C deaminase, APOBEC 3D deaminase, APOBEC 3F deaminase, APOBEC3G deaminase, and APOBEC3H deaminase.

146. The fusion protein of any one of claims 137-145, wherein the cytidine deaminase domain comprises an amino acid sequence that is at least 85%, 90%, 95%, or 98%identical to an amino acid sequence of SEQ ID NO: 49-84.

147. The fusion protein of any one of claims 137-146, wherein the cytidine deaminase domain comprises an amino acid sequence of SEQ ID NO: 49-84.

148. The fusion protein of any one of claims 137-143, wherein the cytidine deaminase domain is an activation-induced deaminase (AID).

149. The fusion protein of any one of claims 137-143, wherein the cytidine deaminase domain is a cytidine deaminase 1 (CDA1).

150. The fusion protein of any one of claims 137-149, wherein the UGI domain comprises a domain capable of inhibiting UDG activity.

151. The fusion protein of any one of claims 137-150, wherein the UGI domain comprises an amino acid sequence that is at least 85%, 90%, 95%, or 98% identical to SEQ ID NO: 134.

152. The fusion protein of any one of claims 137-150, wherein the UGI domain comprises an amino acid sequence as set forth in SEQ ID NO: 134.

153. The fusion protein of any one of claims 137-152, wherein the fusion protein comprises the structure:

NH_2-[cytidine deaminase domain]-[napDNAbp]-[first UGI domain]-[second UGI domain]-COOH;

NH_2-[first UGI domain]-[second UGI domain]-[cytidine deaminase domain]-[napDNAbp]-COOH;

NH_2-[napDNAbp]-[cytidine deaminase domain]-[first UGI domain]-[second UGI domain]-COOH; or

NH_2-[first UGI domain]-[second UGI domain]-[napDNAbp]-[cytidine deaminase domain]-COOH;

wherein each instance of "-" comprises an optional linker.

154. The fusion protein of any one of claims 153, wherein the cytidine deaminase domain and the napDNAbp domain are linked via a linker comprising the amino acid sequence $(GGGS)_n$ (SEQ ID NO: 613), $(GGGGS)_n$ (SEQ ID NO: 607), $(G)_n$ (SEQ ID NO: 608), $(EAAAK)_n$ (SEQ ID NO: 609), $(GGS)_n$ (SEQ ID NO: 610), $(SGGS)_n$ (SEQ ID NO: 606), SGSETPGTSESATPES (SEQ ID NO: 604), SGGS$(GGS)_n$(SEQ ID NO: 612), SGGSSGGSSGS ETPGTS ESATPES SGGSSGGS (SEQ ID NO: 605), or $(XP)_n$ (SEQ ID NO:

611) motif, or a combination thereof, wherein n is independently an integer between 1 and 30, inclusive, and X is any amino acid.

155. The fusion protein of any one of claims 137-154, wherein the cytidine deaminase domain and the napDNAbp are linked via a linker comprising the amino acid sequence: SGGSSGGSSGS ETPGTS ESATPES SGGSSGGS (SEQ ID NO: 605).

156. The fusion protein of any one of claims 137-154, wherein the napDNAbp and the first UGI domain are linked via a linker comprising the amino acid sequence: $SGGS(GGS)_n$ (SEQ ID NO: 612), wherein n is 2.

157. The fusion protein of any one of claims 137-154, wherein the first UGI domain and the second UGI domain are linked via a linker comprising the amino acid sequence: $SGGS(GGS)_n$ (SEQ ID NO: 612), wherein n is 2.

158. The fusion protein of any one of claims 137-157 further comprising a nuclear localization sequence (NLS).

159. The fusion protein of claim 158, wherein the NLS comprises the amino acid sequence PKKKRKV (SEQ ID NO: 614) or KRTADGSEFEPKKKRKV (SEQ ID NO: 740).

160. The fusion protein of claim 158 or 159, wherein the fusion protein comprises the structure:

[N¾]- [cytidine deaminase domain] -[napDNAbp] -[first UGI domain] -[second UGI domain]-[NLS]-[COOH], and each instance of "-" comprises an optional linker.

161. The fusion protein of any one of claims 158-160, wherein the fusion protein comprises the structure:

[cytidine deaminase domain] -[dCas9]- [first UGI domain]-[second UGI domain] -[NLS], and each instance of "-" comprises an optional linker.

162. The fusion protein of any one of claims 160-161, wherein the second UGI domain and the NLS are linked via a linker comprising the amino acid sequence: SGGS (SEQ ID NO: 606).

163. A complex comprising the fusion protein of any one of claims 137-162 and a guide RNA bound to the napDNAbp of the fusion protein.

164. A method comprising contacting a nucleic acid molecule with the fusion protein of any one of claims 137-162 and a guide RNA, wherein the guide RNA comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence in the genome of an organism and comprises a target base pair.

165. The method of claim 164, wherein the target base pair comprises a T to C point mutation associated with a disease or disorder, and wherein the deamination of the mutant C base results in a sequence that is not associated with a disease or disorder.

166. The method of claim 164 or 165, wherein the contacting results in less than 20% indel formation upon base editing.

167. The method of claim 164 or 165, wherein the target sequence comprises a point mutation associated with a disease or disorder.

168. The method of any one of claims 164-167, wherein the activity of the fusion protein, or the complex results in a correction of the point mutation.

169. The method of any one of claims 164-165, wherein the target sequence comprises a T to C point mutation associated with a disease or disorder, and wherein the deamination of the mutant C base results in a sequence that is not associated with a disease or disorder.

170. The method of claim 169, wherein the target sequence encodes a protein, and wherein the point mutation is in a codon and results in a change in the amino acid encoded by the mutant codon as compared to a wild-type codon.

171. The method of claim 169 or 170, wherein the deamination of the mutant C results in a change of the amino acid encoded by the mutant codon.

172. The method of claim 169 or 170, wherein the deamination of the mutant C results in the codon encoding a wild-type amino acid.

173. The method of any one of claims 164-172, wherein the contacting is performed *in vivo* in a subject.

174. The method of any one of claims 164-172, wherein the contacting is performed *in vitro.*

175. The method of claim 173, wherein the subject has been diagnosed with a disease or disorder.

176. A method for editing a nucleobase pair of a double-stranded DNA sequence, the method comprising:

contacting a target region of the double-stranded DNA sequence with a complex comprising a nucleobase editor and a guide nucleic acid, wherein the target region comprises a target nucleobase pair;

inducing strand separation of said target region;

converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase; and

cutting no more than one strand of said target region;

wherein a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase;

wherein the method causes less than 20% indel formation in the double-stranded DNA sequence; and

wherein the nucleobase editor comprises CasX, CasY, Cpfl, C2cl, C2c2, C2c3, or Argonaute.

177. The method of claim 176, wherein the method causes less than 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% indel formation.

178. The method of claim 176, further comprising replacing the second nucleobase with a fifth nucleobase that is complementary to the fourth nucleobase, thereby generating an intended edited base pair.

179. The method of claim 178, wherein the efficiency of generating the intended edited base pair is at least 5%.

180. The method of claim 179, wherein the efficiency is at least 10%, 20%, 30%, 40%, or 50%.

181. The method of claim 178, wherein the ratio of intended edited basepairs to unintended edited basepairs is 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, or 8:1.

182. The method of claim 178, wherein the ratio of intended edited basepairs to indel formation is from 2:1 through 1000:1.

183. The method of claim 176, wherein the cut single strand is hybridized to the guide nucleic acid.

184. The method of claim 176, wherein the cut single strand is opposite to the strand comprising the first nucleobase.

185. The method of claim 176, wherein said first base is cytosine.

186. The method of claim 176, wherein the second nucleobase is a non-standard nucleobase.

187. The method of claim 176, wherein the second nucleobase is other than G, C, A, or T.

188. The method of claim 176, wherein said second base is uracil.

189. The method of claim 176, wherein the nucleobase editor comprises UGI activity.

190. The method of claim 176, wherein the nucleobase editor comprises a UGI domain.

191. The method of claim 176, wherein the nucleobase editor comprises two UGI domains.

192. The method of claim 176, wherein the nucleobase editor comprises nickase activity.

193. The method of claim 178, wherein the intended edited base pair is upstream of a PAM site.

194. The method of claim 193, wherein the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream of the PAM site.

195. The method of claim 178, wherein the intended edited base pair is downstream of a PAM site.

196. The method of claim 195, wherein the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides downstream stream of the PAM site.

197. The method of claim 176, wherein the method does not require a canonical PAM site.

198. The method of claim 176, wherein the nucleobase editor comprises a linker.

199. The method of claim 198, wherein the linker is 1-25 amino acids in length.

200. The method of claim 198, wherein the linker is 5-20 amino acids in length.

201. The method of claim 198, wherein the linker is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length.

202. The method of claim 178, wherein the target region comprises a target window, wherein the target window comprises the target nucleobase pair.

203. The method of claim 202, wherein the target window comprises 1-10 nucleotides.

204. The method of claim 202, wherein the target window is 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2, or 1 nucleotides in length.

205. The method of claim 202, wherein the target window is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length.

206. The method of claim 202-205, wherein the intended edited base pair occurs within the target window.

207. The method of claim 202-206, wherein the target window comprises the intended edited base pair.

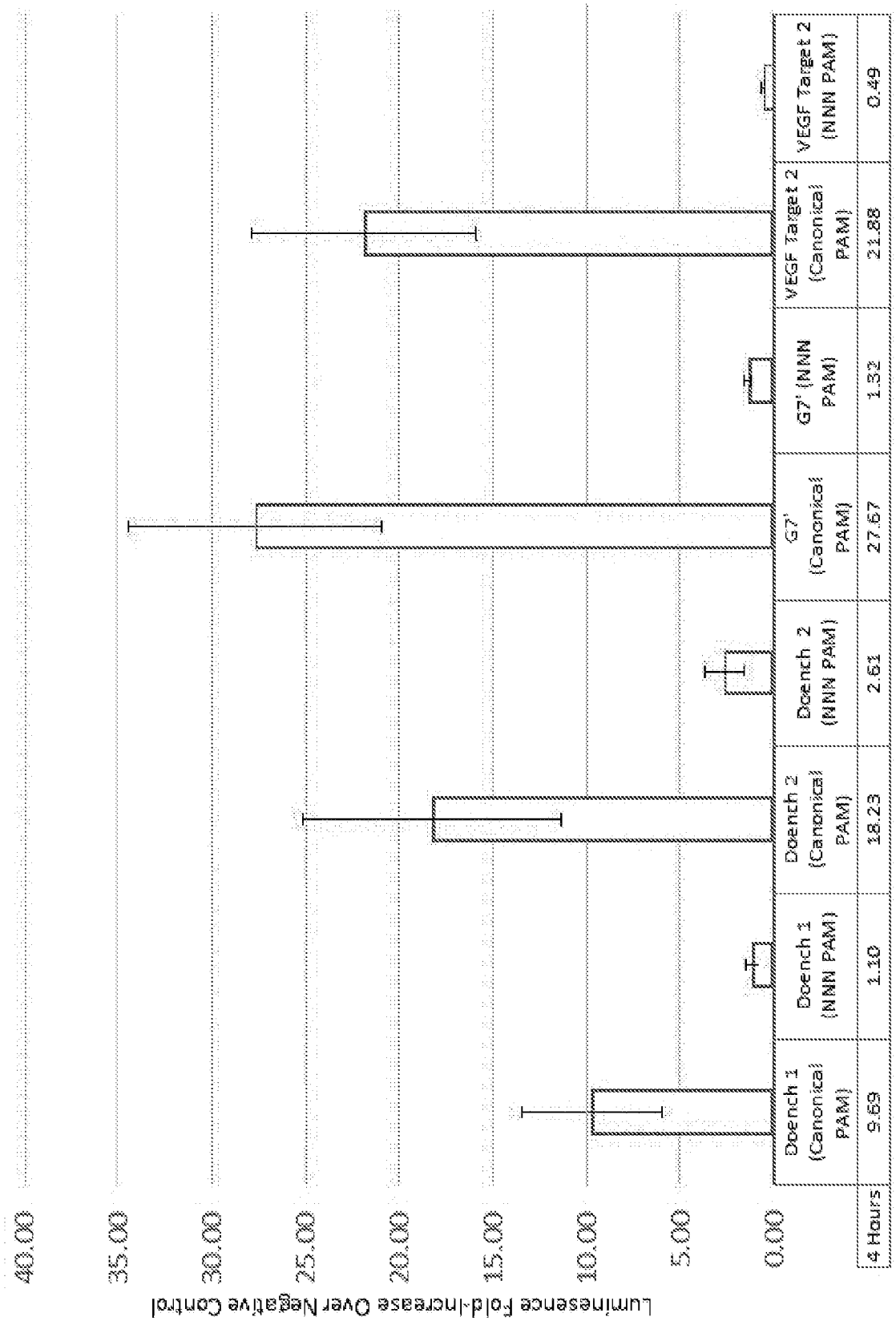

Luminescence Fold-Increase Over Negative Control
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
4 Hours
Doench 1 (Canonical PAM)
9.69
Doench 1 (NNN PAM)
1.10
Doench 2 (Canonical PAM)
18.23
Doench 2 (NNN PAM)
2.61
G7' (Canonical PAM)
27.67
G7' (NNN PAM)
1.32
VEGF Target 2 (Canonical PAM)
21.88
VEGF Target 2 (NNN PAM)
0.49

FIGURE 1

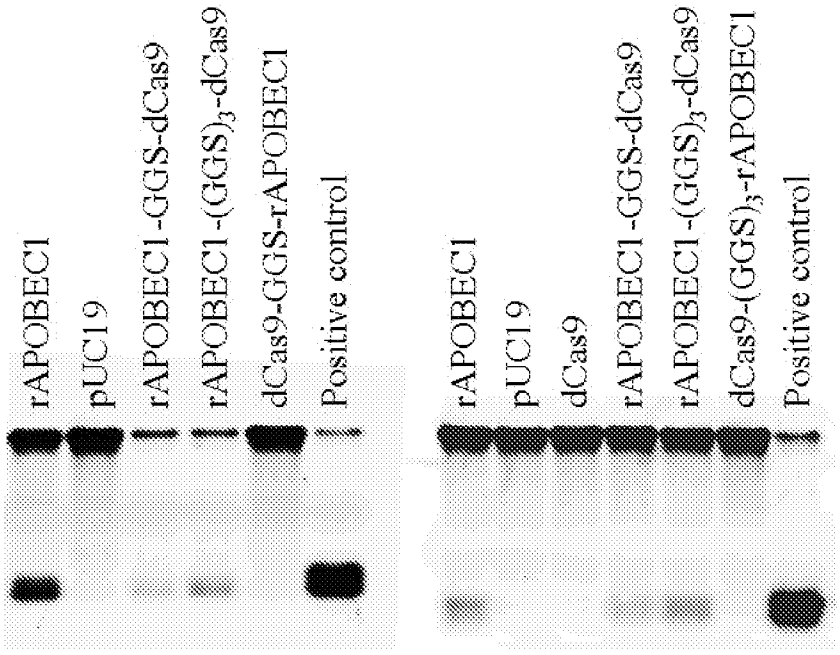

rAPOBEC1
pUC19
rAPOBEC1-GGS-dCas9
rAPOBEC1-(GGS)3-dCas9
dCas9-GGS-rAPOBEC1
Positive control
rAPOBEC1
pUC19
dCas9
rAPOBEC1-GGS-dCas9
rAPOBEC1-(GGS)3-dCas9
dCas9-(GGS)3-rAPOBEC1
Positive control

FIGURE 2

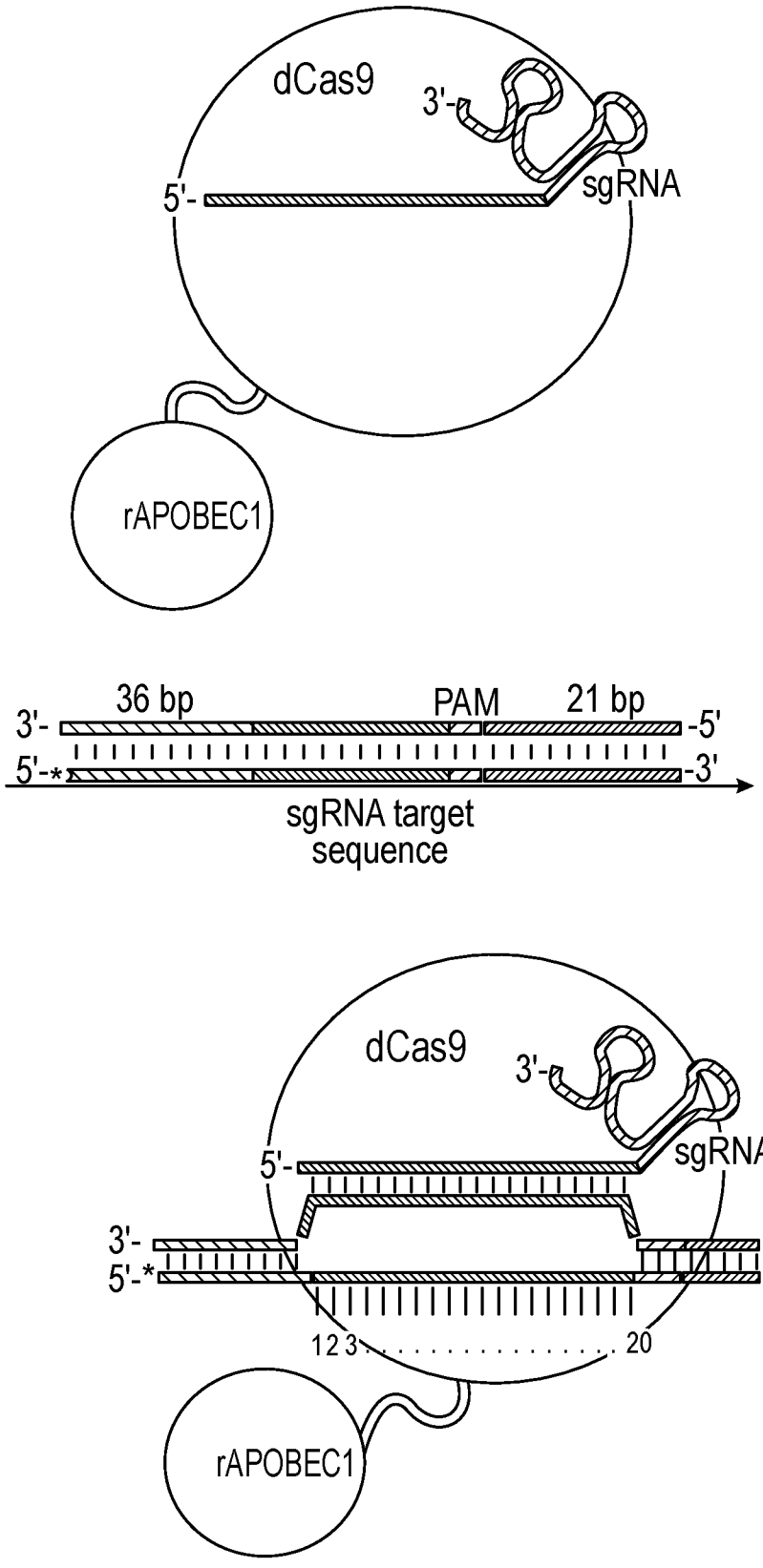

dCas9
3'-
5'-
sgRNA
rAPOBEC1
36 bp
PAM
21 bp
3'-
-5'
5'-*
-3'
sgRNA target sequence
dCas9
3'-
5'-
sgRNA
3'-
-5'
5'-*
-3'
1 2 3 20
rAPOBEC1

FIGURE 3

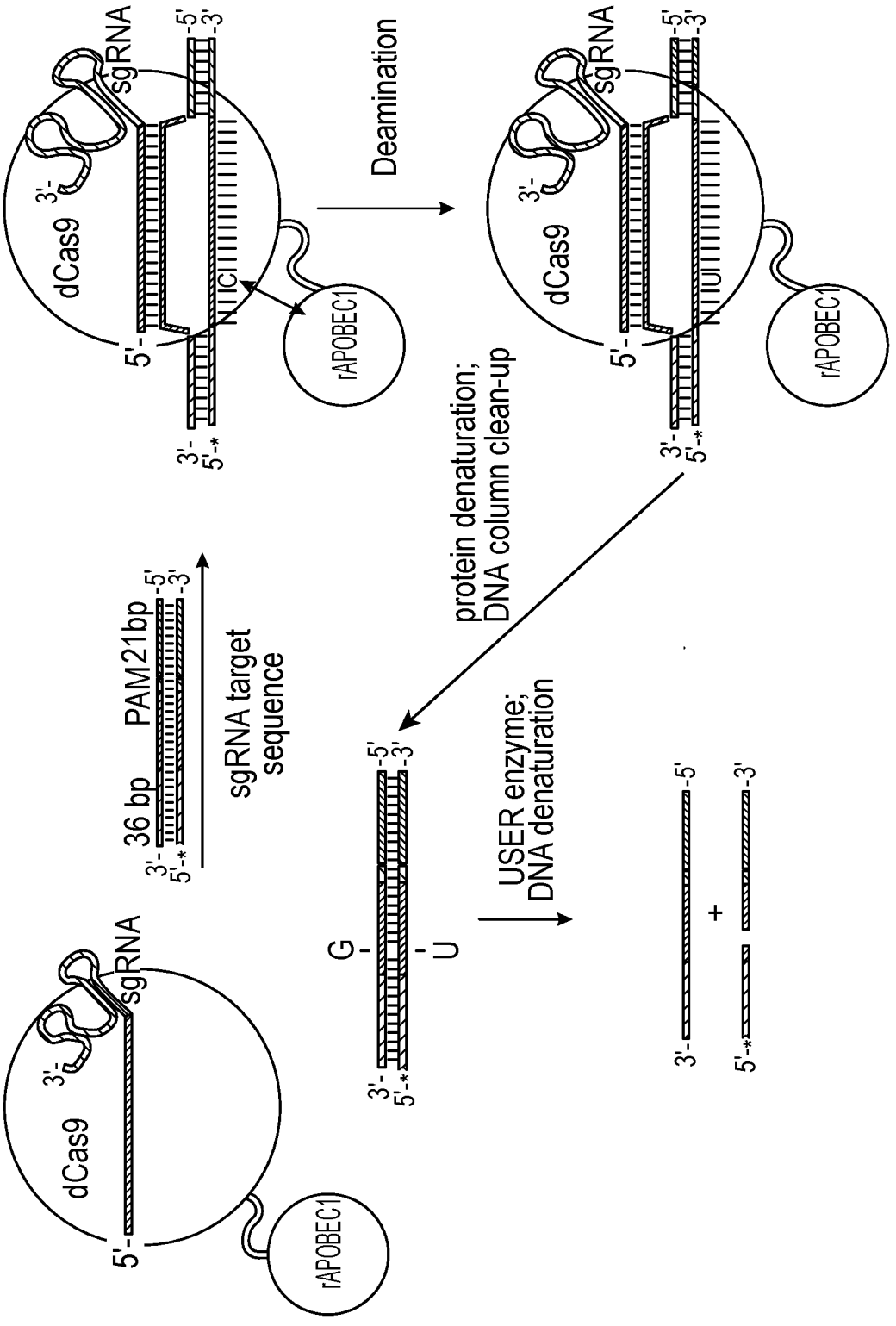

dCas9
sgRNA
rAPOBEC1
Deamination
protein denaturation;
DNA column clean-up
USER enzyme;
DNA denaturation
sgRNA target
sequence
36 bp
PAM21bp

FIGURE 4

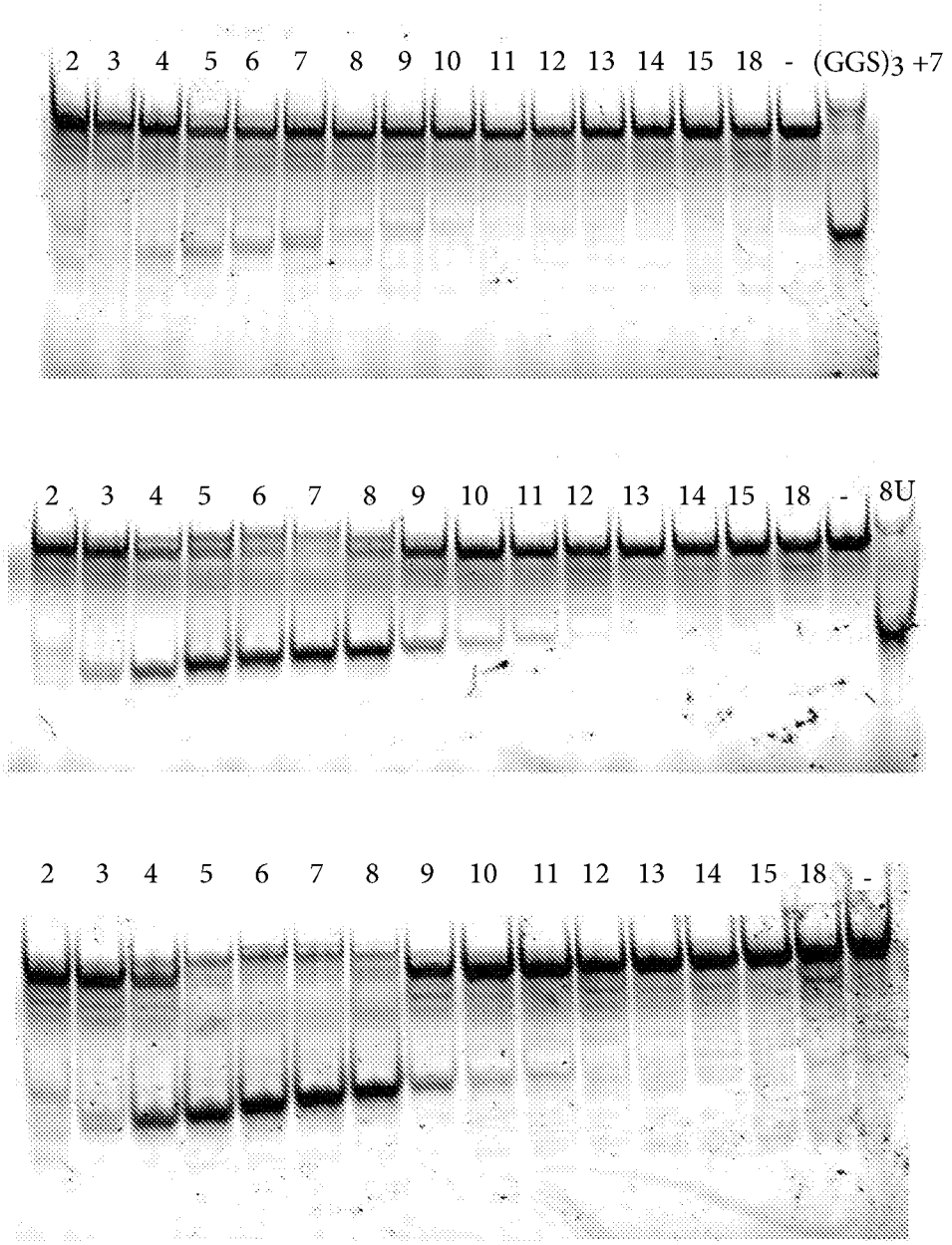

2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 - (GGS)3 +7
2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 - 8U
2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 -

FIGURE 5

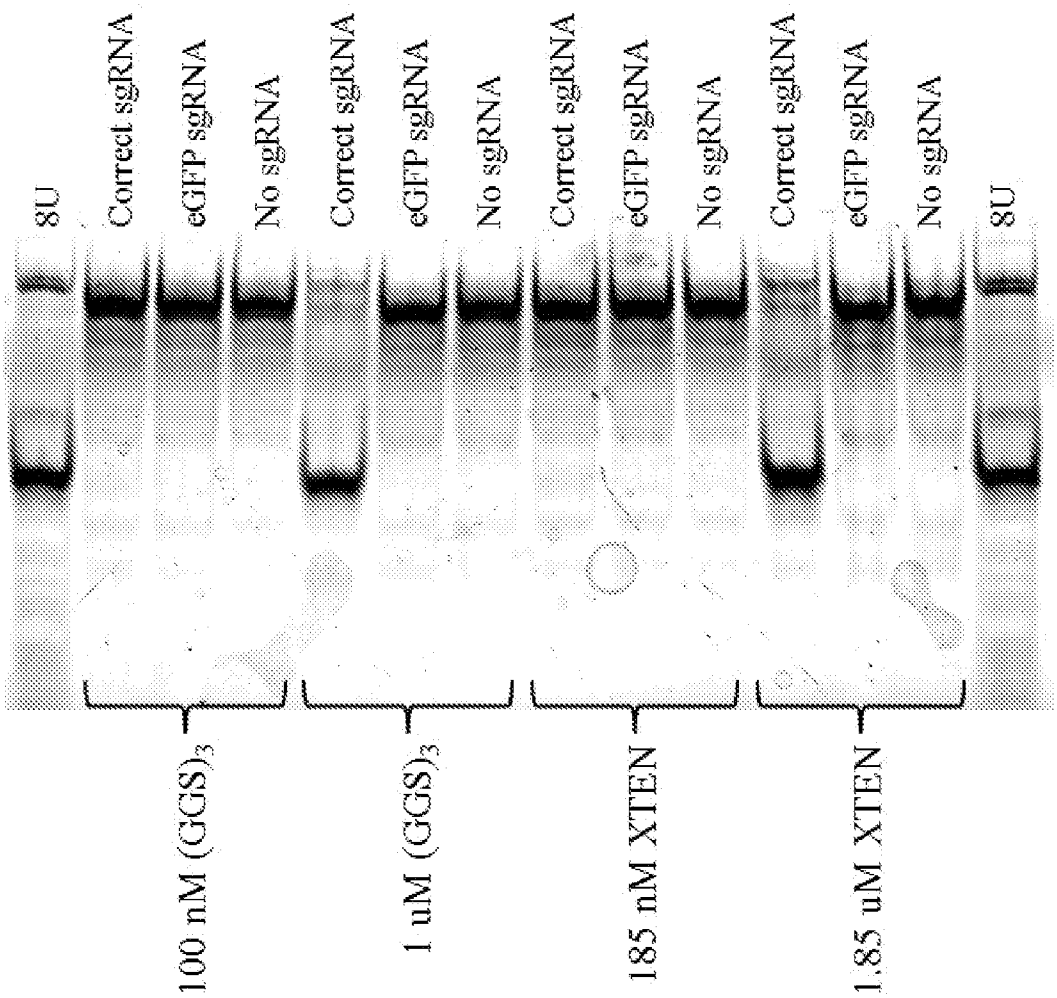

8U
Correct sgRNA
eGFP sgRNA
No sgRNA
Correct sgRNA
eGFP sgRNA
No sgRNA
Correct sgRNA
eGFP sgRNA
No sgRNA
Correct sgRNA
eGFP sgRNA
No sgRNA
8U
100 nM (GGS)3
1 uM (GGS)3
185 nM XTEN
1.85 uM XTEN

FIGURE 6

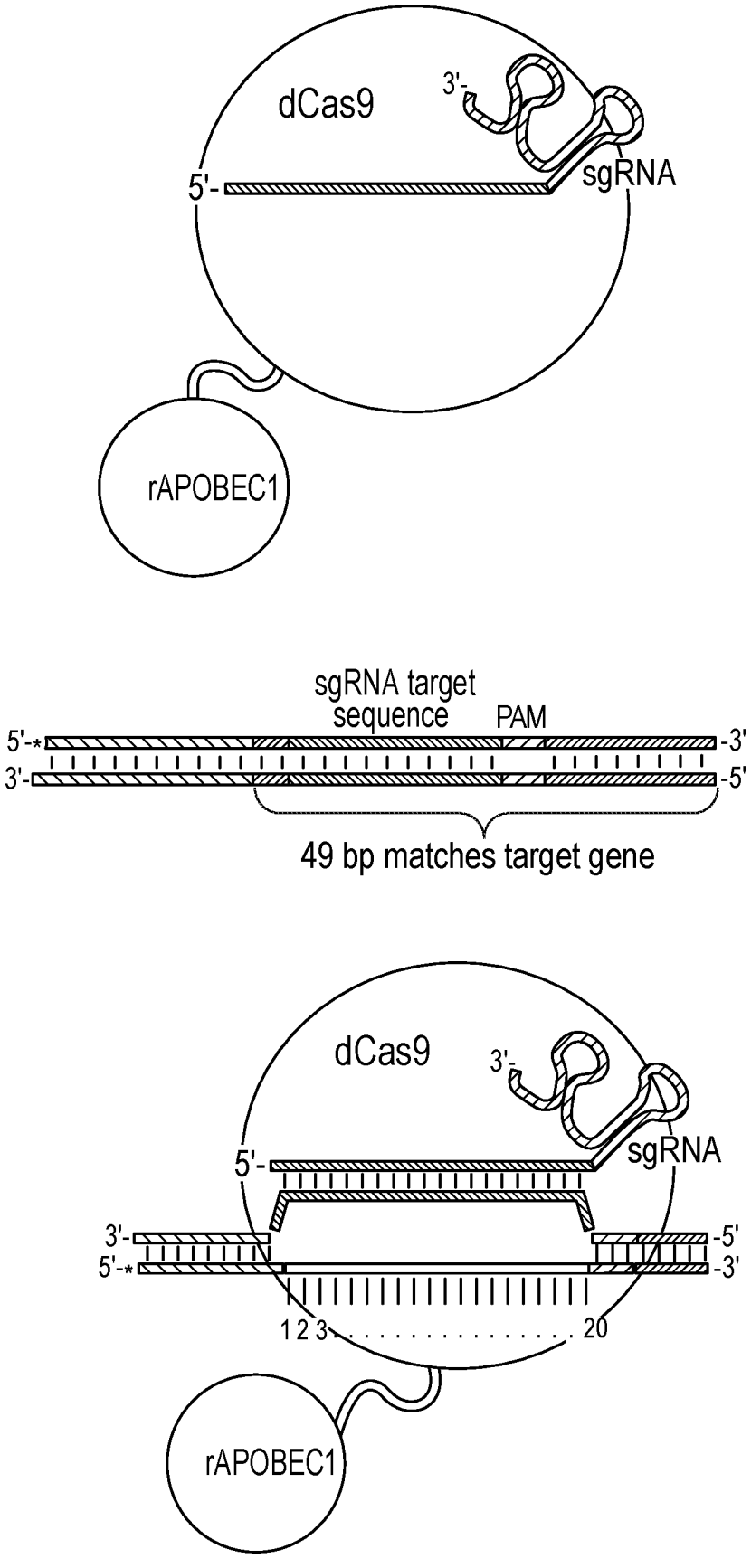

dCas9
3'-
5'-
sgRNA
rAPOBEC1
sgRNA target
sequence
PAM
5'-*
-3'
3'-
-5'
49 bp matches target gene
dCas9
3'-
5'-
sgRNA
3'-
-5'
5'-*
-3'
1 2 3 20
rAPOBEC1

FIGURE 7

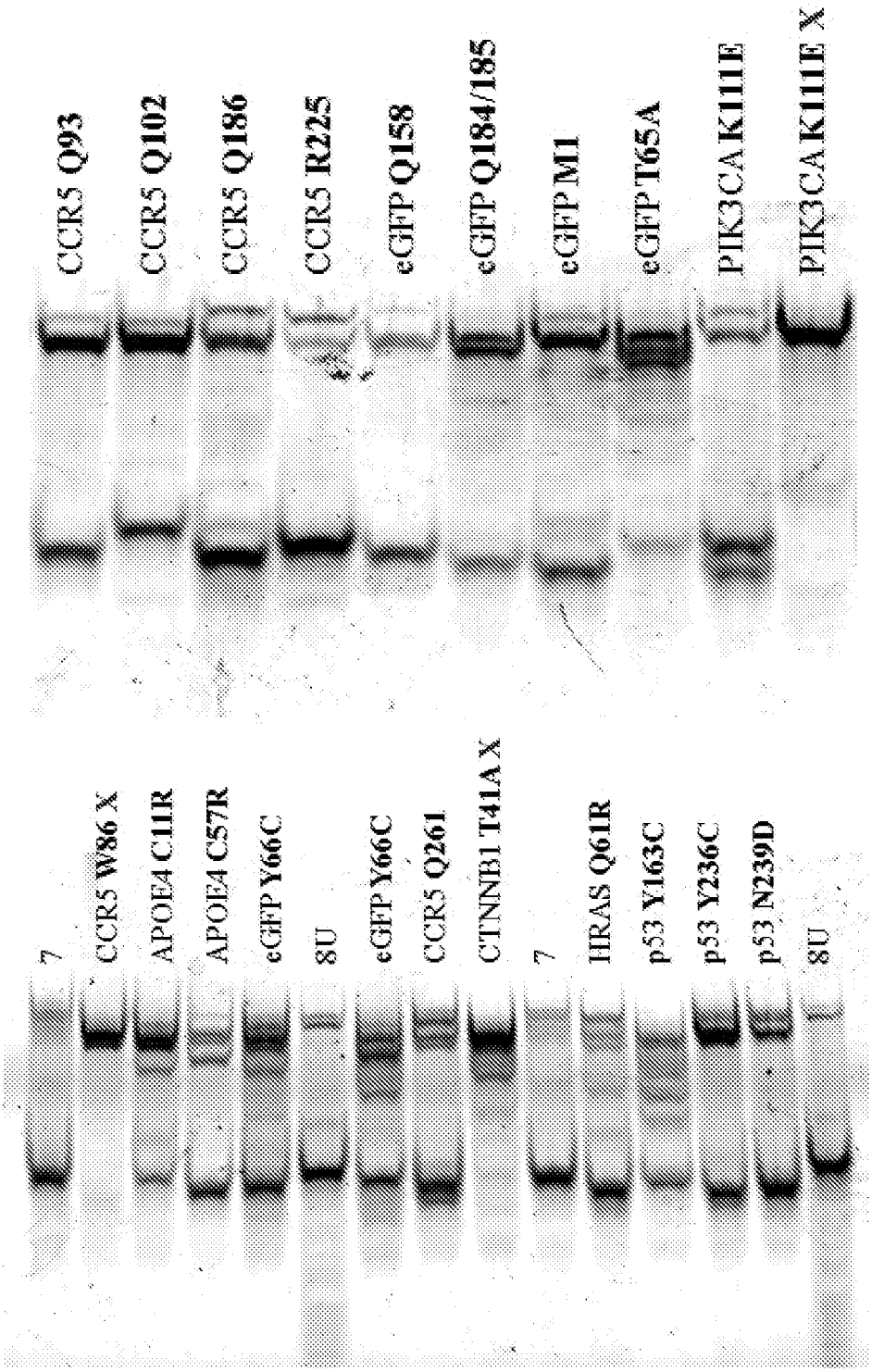

7
CCR5 W86 X
APOE4 C11R
APOE4 C57R
eGFP Y66C
8U
eGFP Y66C
CCR5 Q261
CTNNB1 T41A X
7
HRAS Q61R
p53 Y163C
p53 Y236C
p53 N239D
8U
CCR5 Q93
CCR5 Q102
CCR5 Q186
CCR5 R225
eGFP Q158
eGFP Q184/185
eGFP M1
eGFP T65A
PIK3CA K111E
PIK3CA K111E X

FIGURE 8

EMX1		C_5	C_6	C_{10}
0 uM XTEN	A	0.1%	0.1%	0.1%
	C	99.8%	99.8%	99.8%
	G	0.0%	0.0%	0.1%
	T	0.0%	0.1%	0.0%
1.85 uM XTEN	A	0.1%	0.0%	0.1%
	C	60.4%	61.0%	99.1%
	G	0.0%	0.0%	0.1%
	T	39.5%	39.0%	0.7%

FANCF		C_6	C_7	C_8	C_{11}
0 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	99.8%	99.8%	99.9%	99.9%
	G	0.0%	0.1%	0.0%	0.0%
	T	0.0%	0.0%	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	63.9%	64.7%	65.0%	72.6%
	G	0.0%	0.0%	0.0%	0.0%
	T	36.0%	35.1%	34.9%	27.3%

HEK293 site 2		C_4	C_6	C_{11}
0 uM XTEN	A	0.1%	0.1%	0.1%
	C	99.9%	99.9%	99.9%
	G	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.1%
1.85 uM XTEN	A	0.1%	0.1%	0.1%
	C	80.6%	76.9%	99.6%
	G	0.0%	0.0%	0.0%
	T	19.3%	22.9%	0.3%

HEK293 site 3		C_3	C_4	C_5	C_9
0 uM XTEN	A	0.1%	0.1%	0.0%	0.1%
	C	99.8%	99.9%	99.9%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.1%	0.0%	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.0%	0.1%
	C	92.2%	74.8%	71.5%	96.6%
	G	0.0%	0.0%	0.0%	0.0%
	T	7.7%	25.1%	28.5%	3.3%

HEK293 site 4		C_3	C_5	C_8	C_{11}
0 uM XTEN	A	0.1%	0.0%	0.1%	0.0%
	C	99.8%	99.9%	99.8%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.1%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	98.8%	60.1%	97.0%	99.4%
	G	0.0%	0.0%	0.0%	0.0%
	T	1.1%	39.8%	2.9%	0.5%

RNF2		C_3	C_6
0 uM XTEN	A	0.1%	0.0%
	C	99.9%	99.9%
	G	0.0%	0.0%
	T	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.0%
	C	59.1%	57.8%
	G	0.0%	0.0%
	T	40.8%	42.1%

FIGURE 9

EMX1		C_5	C_6	C_{10}
untreated	A	0.0%	0.0%	0.0%
	C	99.5%	99.7%	100.0%
	G	0.0%	0.1%	0.0%
	T	0.5%	0.2%	0.0%
XTEN	A	0.7%	0.5%	0.0%
	C	93.5%	95.8%	100.0%
	G	2.1%	0.3%	0.0%
	T	3.6%	3.3%	0.0%
XTEN-UGI	A	0.2%	0.0%	0.0%
	C	81.8%	82.5%	100.0%
	G	0.6%	0.3%	0.0%
	T	17.4%	17.1%	0.0%

FANCF		C_6	C_7	C_8	C_{11}
untreated	A	0.0%	0.0%	0.2%	0.1%
	C	99.9%	99.8%	99.8%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.1%	0.1%	0.0%	0.0%
XTEN	A	0.3%	0.1%	0.0%	0.0%
	C	98.1%	99.2%	99.0%	99.8%
	G	0.4%	0.0%	0.0%	0.0%
	T	1.2%	0.7%	1.0%	0.2%
XTEN-UGI	A	0.0%	0.0%	0.1%	0.0%
	C	93.2%	93.5%	93.4%	98.2%
	G	0.0%	0.0%	0.0%	0.0%
	T	6.7%	6.5%	6.5%	1.8%

HEK293 site 2		C_4	C_6	C_{11}
untreated	A	0.3%	0.2%	0.2%
	C	99.7%	99.7%	99.7%
	G	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.0%
XTEN	A	0.3%	0.3%	0.3%
	C	99.7%	99.4%	99.7%
	G	0.0%	0.3%	0.0%
	T	0.0%	0.0%	0.0%
XTEN-UGI	A	0.3%	0.2%	0.2%
	C	98.8%	98.2%	99.8%
	G	0.0%	0.3%	0.0%
	T	0.9%	1.3%	0.0%

HEK293 site 2		C_3	C_4	C_5	C_9
untreated	A	0.0%	0.0%	0.0%	0.0%
	C	100.0%	100.0%	100.0%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.0%	0.1%
XTEN	A	0.0%	0.6%	0.3%	0.1%
	C	100.0%	95.8%	95.8%	99.2%
	G	0.0%	0.2%	0.7%	0.4%
	T	0.0%	3.4%	3.2%	0.3%
XTEN-UGI	A	0.0%	0.3%	0.3%	0.0%
	C	96.8%	83.0%	79.2%	98.5%
	G	0.0%	0.0%	1.1%	0.2%
	T	3.2%	16.8%	19.4%	1.3%

HEK293 site 4		C_3	C_5	C_8	C_{11}
untreated	A	0.0%	0.4%	0.0%	0.0%
	C	99.8%	97.6%	99.9%	100.0%
	G	0.0%	1.0%	0.0%	0.0%
	T	0.2%	1.0%	0.0%	0.0%
XTEN	A	0.0%	1.1%	0.0%	0.0%
	C	99.6%	92.2%	99.9%	100.0%
	G	0.0%	2.2%	0.0%	0.0%
	T	0.4%	4.5%	0.0%	0.0%
XTEN-UGI	A	0.0%	0.5%	0.0%	0.0%
	C	99.4%	86.7%	99.1%	100.0%
	G	0.0%	1.8%	0.0%	0.0%
	T	0.6%	11.0%	0.9%	0.0%

RNF2		C_3	C_6
untreated	A	0.0%	0.0%
	C	99.9%	99.5%
	G	0.0%	0.2%
	T	0.0%	0.3%
XTEN	A	0.0%	0.0%
	C	99.8%	99.3%
	G	0.0%	0.2%
	T	0.2%	0.5%
XTEN-UGI	A	0.0%	0.0%
	C	99.6%	99.1%
	G	0.0%	0.4%
	T	0.4%	0.5%

FIGURE 10

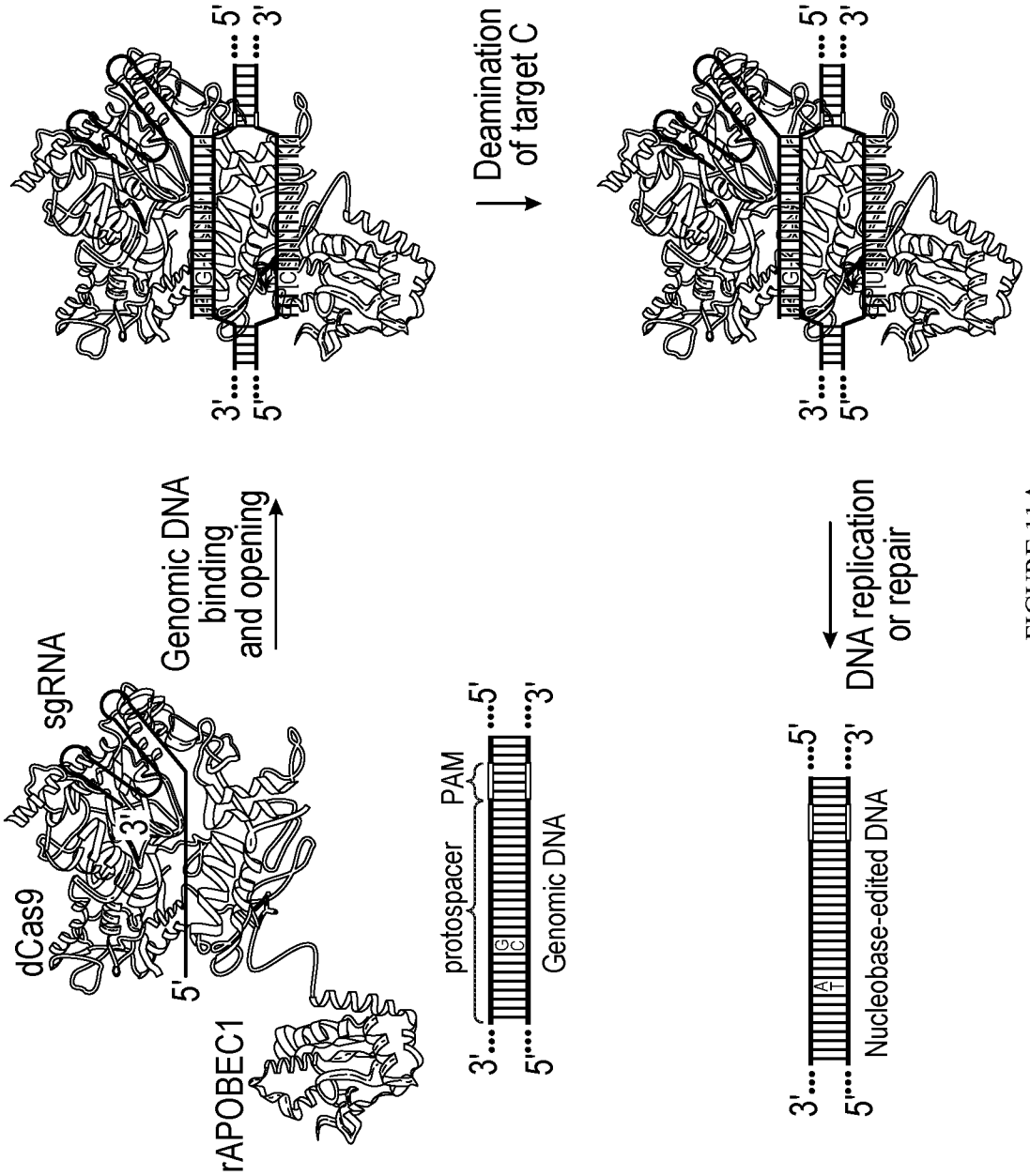

dCas9
sgRNA
3'
5'
rAPOBEC1
Genomic DNA
binding
and opening
protospacer
PAM
3'
5'
C
G
Genomic DNA
Deamination
of target C
U
A
Nucleobase-edited DNA
DNA replication
or repair

FIGURE 11A

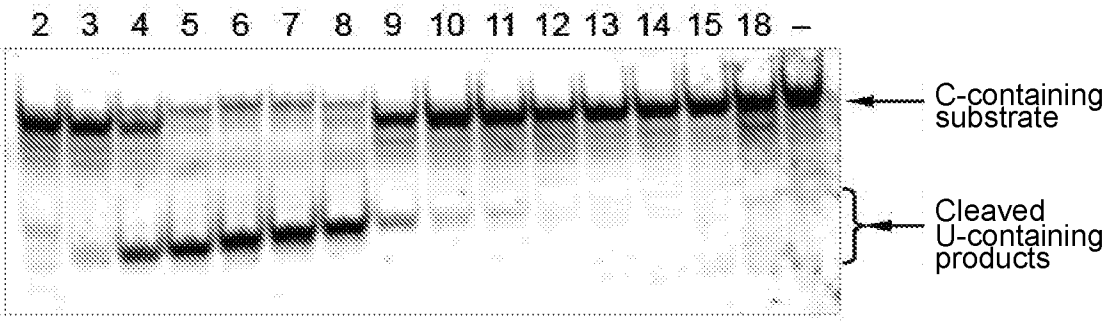


FIGURE 11B

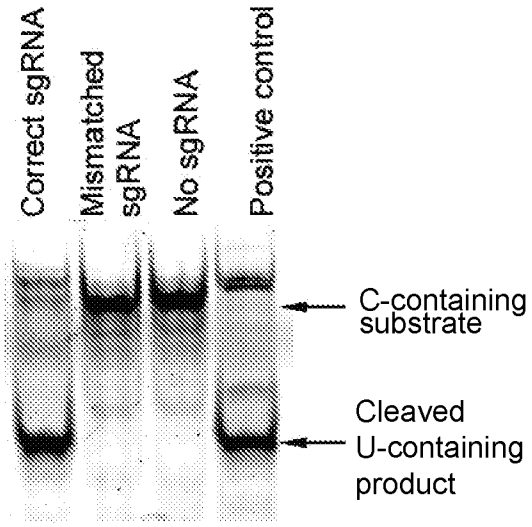


FIGURE 11C

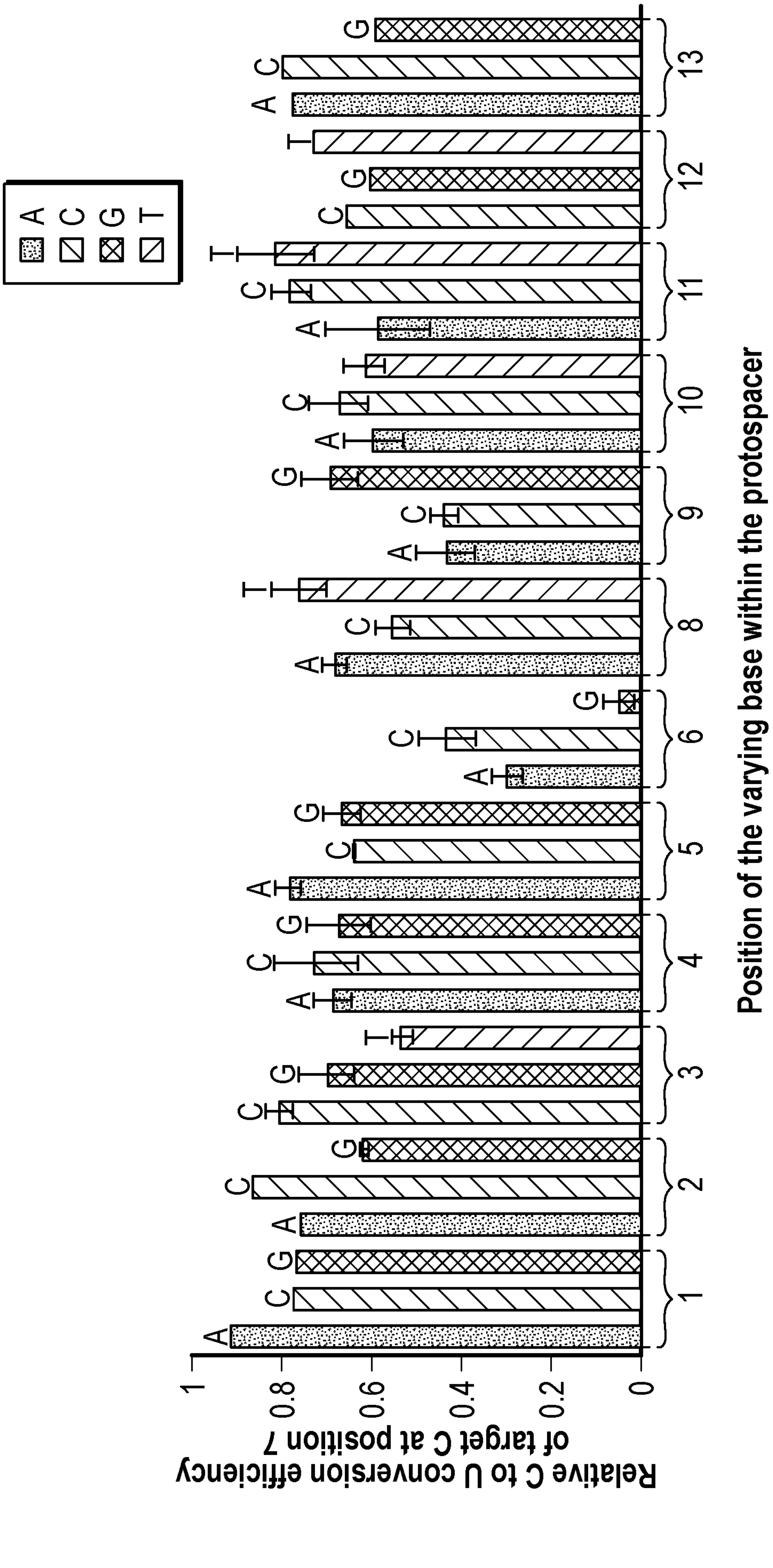

A
C
G
T
Relative C to U conversion efficiency of target C at position 7
1
0.8
0.6
0.4
0.2
0
Position of the varying base within the protospacer
1
2
3
4
5
6
8
9
10
11
12
13

FIGURE 12A

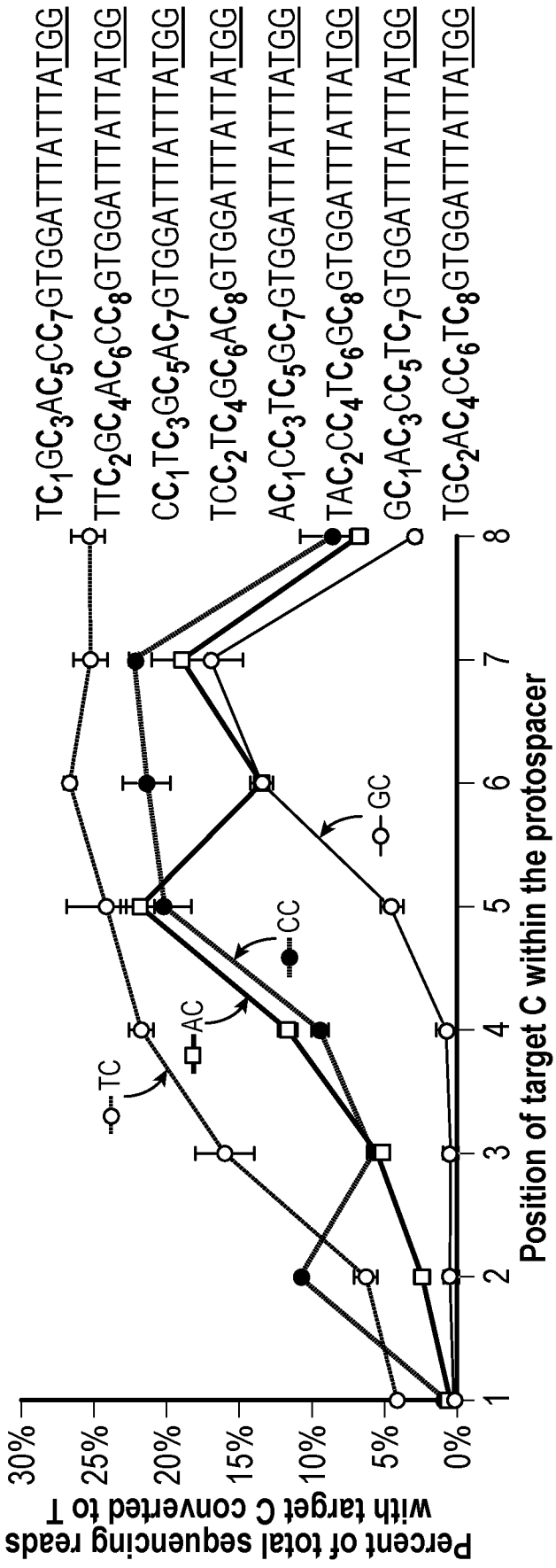

Percent of total sequencing reads with target C converted to T
30%
25%
20%
15%
10%
5%
0%
1
2
3
4
5
6
7
8
Position of target C within the protospacer
TC
AC
CC
GC
TC1GC3AC5CC7GTGGATTTATTTATGG
TTC2GC4AC6CC8GTGGATTTATTTATGG
CC1TC3GC5AC7GTGGATTTATTTATGG
TCC2TC4GC6AC8GTGGATTTATTTATGG
AC1CC3TC5GC7GTGGATTTATTTATGG
TAC2CC4TC6GC8GTGGATTTATTTATGG
GC1AC3CC5TC7GTGGATTTATTTATGG
TGC2AC4CC6TC8GTGGATTTATTTATGG

FIGURE 12B

EMX1:	GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG
FANCF:	GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG
HEK293 site 2:	GAAC$_4$AC$_6$AAAGCATAGACTGCGGG
HEK293 site 3:	GGCC$_4$C$_5$AGACTGAGCACGTGATGG
HEK293 site 4:	GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2:	GTC$_3$ATC$_6$TTAGTC$_{12}$ATTACCTGAGG

FIGUIRE 13A

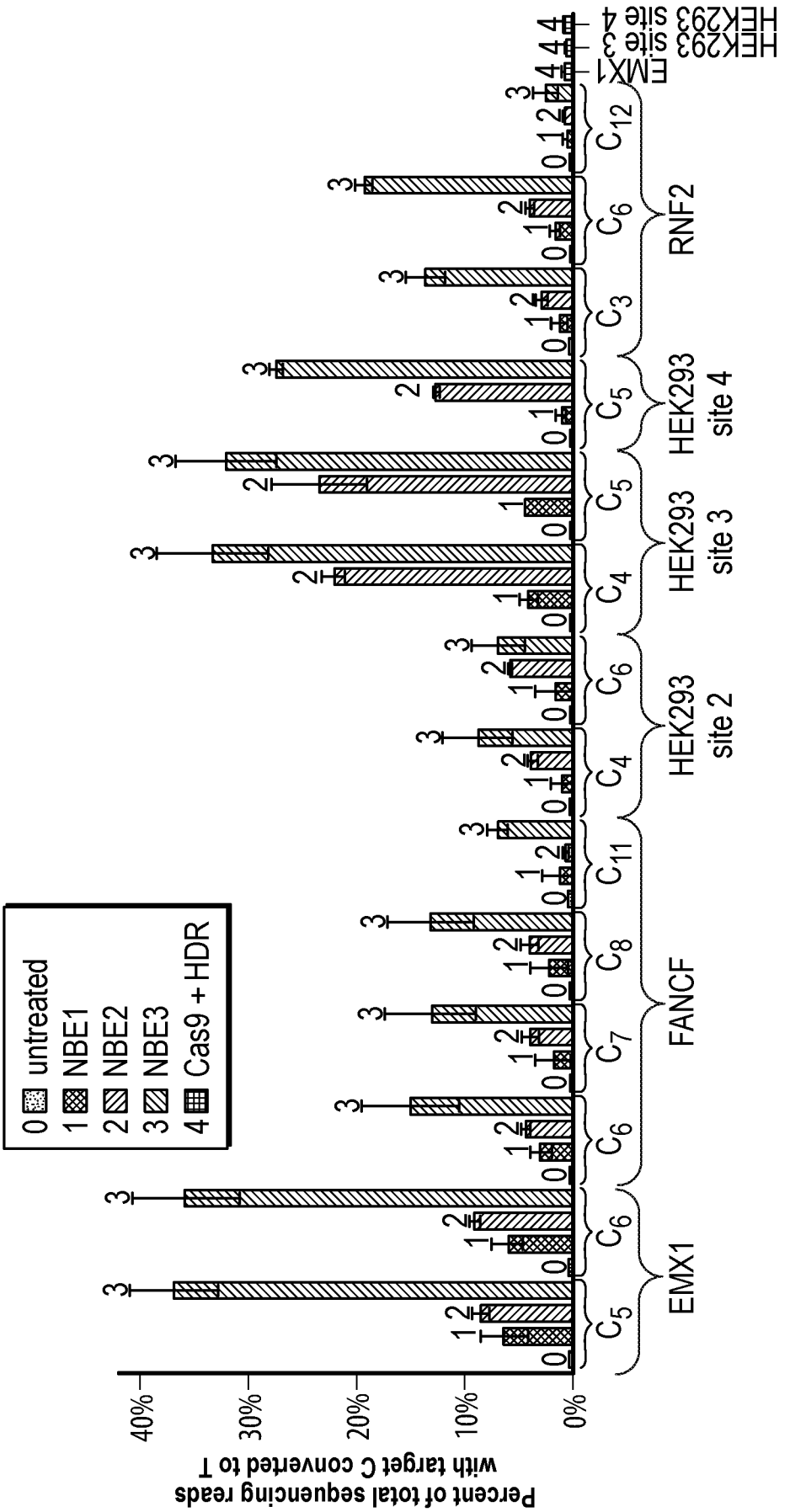

Percent of total sequencing reads with target C converted to T
40%
30%
20%
10%
0%
0 untreated
1 NBE1
2 NBE2
3 NBE3
4 Cas9 + HDR
EMX1
FANCF
HEK293 site 2
HEK293 site 3
HEK293 site 4
RNF2
EMX1
HEK293 site 3
HEK293 site 4

FIGURE 13B

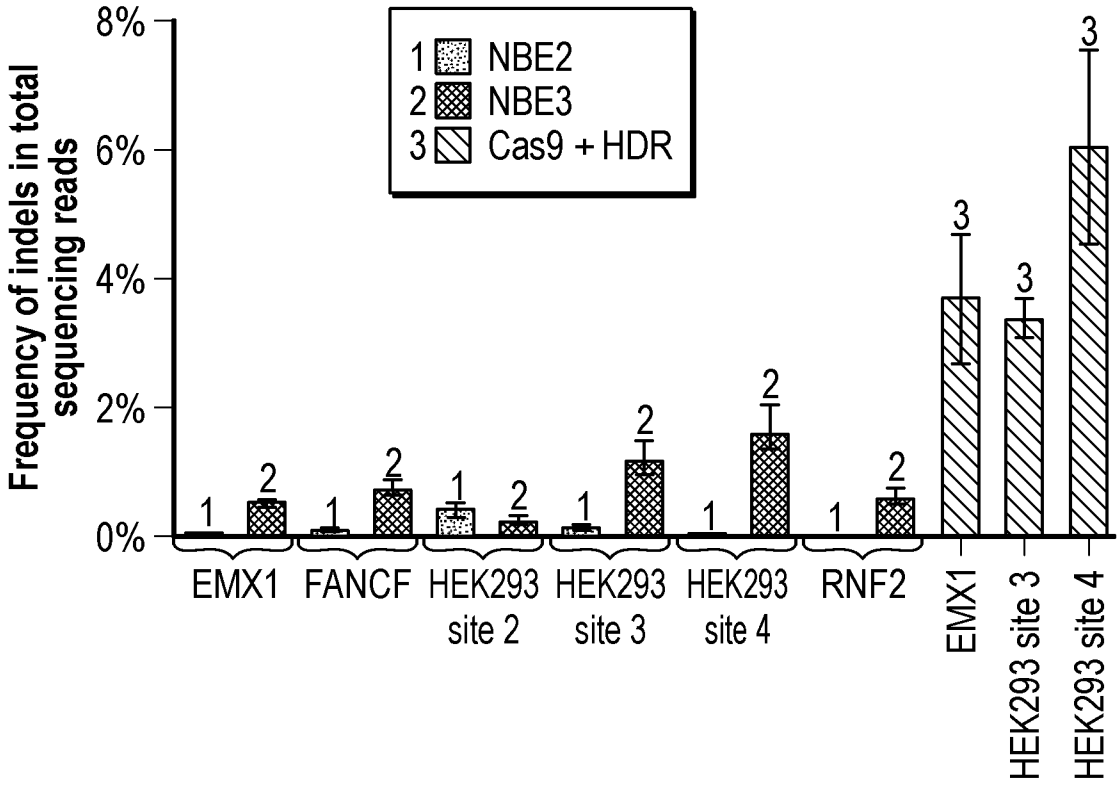

Frequency of indels in total sequencing reads
8%
6%
4%
2%
0%
1 NBE2
2 NBE3
3 Cas9 + HDR
EMX1
FANCF
HEK293 site 2
HEK293 site 3
HEK293 site 4
RNF2
EMX1
HEK293 site 3
HEK293 site 4

FIGURE 13C

		Lys			Arg → Cys			Leu → Leu			Ala			Val			Tyr			Gln			
APOE4 C158R	G	A	A	G	**C_5**	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G
A	0.0	100.0	100.0	0.0	0.2	0.0	0.1	0.3	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0
C	0.0	0.0	0.0	0.0	99.7	0.0	91.3	91.1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0
G	100.0	0.0	0.0	100.0	0.2	100.0	0.4	0.1	0.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0
T	0.0	0.0	0.0	0.0	10.9	0.0	8.2	8.4	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0

FIGURE 14A

		His			Tyr			Asn			Tyr			Met			Cys			Asp → Asn			
TP53 N239D	C	C	A	C	T	A	C	A	A	C	T	A	C	A	T	G	T	G	T	**G_4**	A	C	A
A	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	55.7	100.0	0.0	100.0
C	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	44.3	0.0	0.0	0.0
T	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0

FIGURE 14B

		Arg			Ala			Met			Ala			Ile			Cys → Tyr			Lys			
TP53 Y163C	C	C	G	C	G	C	C	A	T	G	G	C	C	A	T	C	T	**G_6**	C	A	A	G	C
A	0.0	0.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	7.6	0.0	100.0	100.0	0.0	0.0
C	100.0	100.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	99.9	100.0	0.0	0.0	100.0	0.0	0.4	100.0	0.0	0.0	0.0	100.0
G	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0	0.0	91.8	0.0	0.0	0.0	100.0	0.0
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	0.1	0.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0

FIGURE 14C

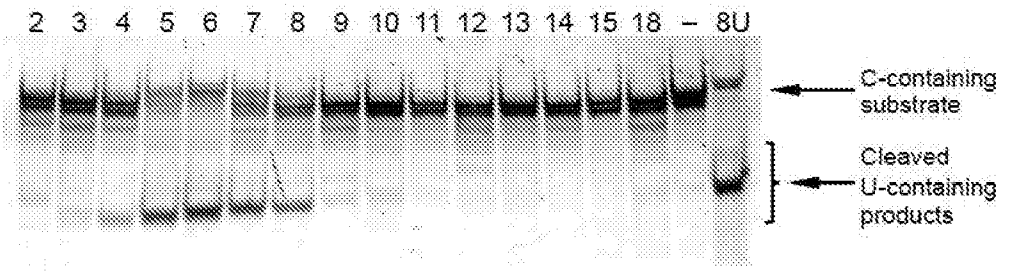


FIGURE 15A

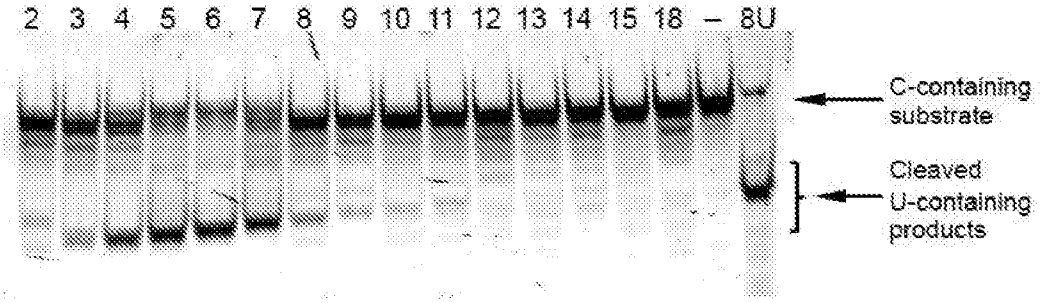


FIGURE 15B

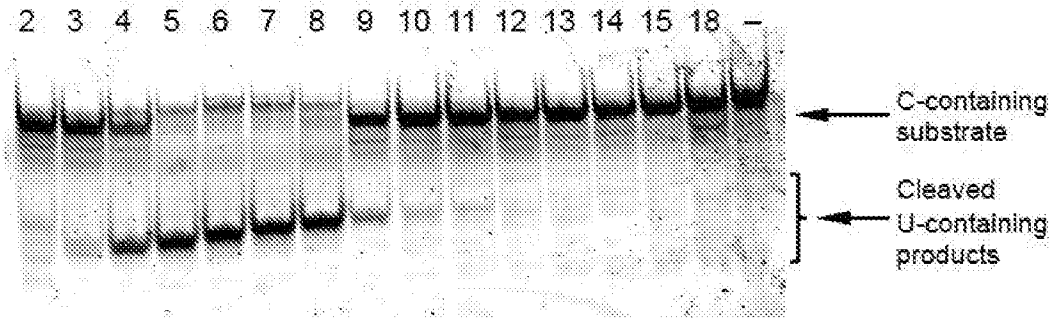


FIGURE 15C

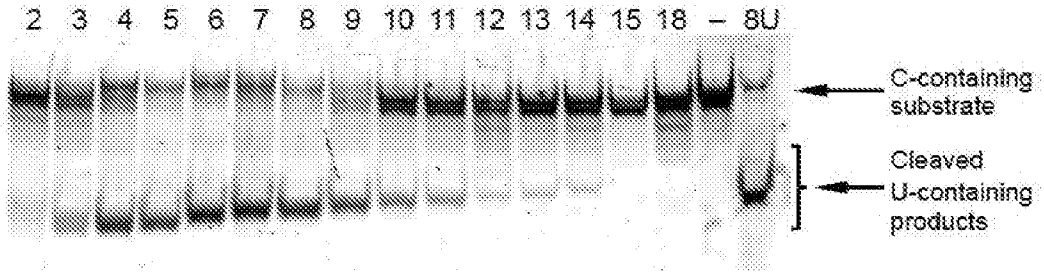


FIGURE 15D

APOE4 Cys112Arg: 5'-GGAGGACGTGC$_{11}$GCGGCCGCCTGG
APOE4 Cys158Arg: 5'-GAAGC$_5$GCCTGGCAGTGTACCAGG
CTNNB1 Thr41Ala: 5'-CTGTGGC$_7$AGTGGCACCAGAATGG
HRAS Gln61Arg: 5'-CCTCCC$_6$GGCCGGCGGTATCCAGG
p53 Tyr163Cys: 5'-GCTTGC$_6$AGATGGCCATGGCGCGG
p53 Tyr236Cys: 5'-ACACATGC$_8$AGTTGTAGTGGATGG
p53 Asn239Asp: 5'-TGTC$_4$ACACATGTAGTTGTAGTGG

FIGURE 16A

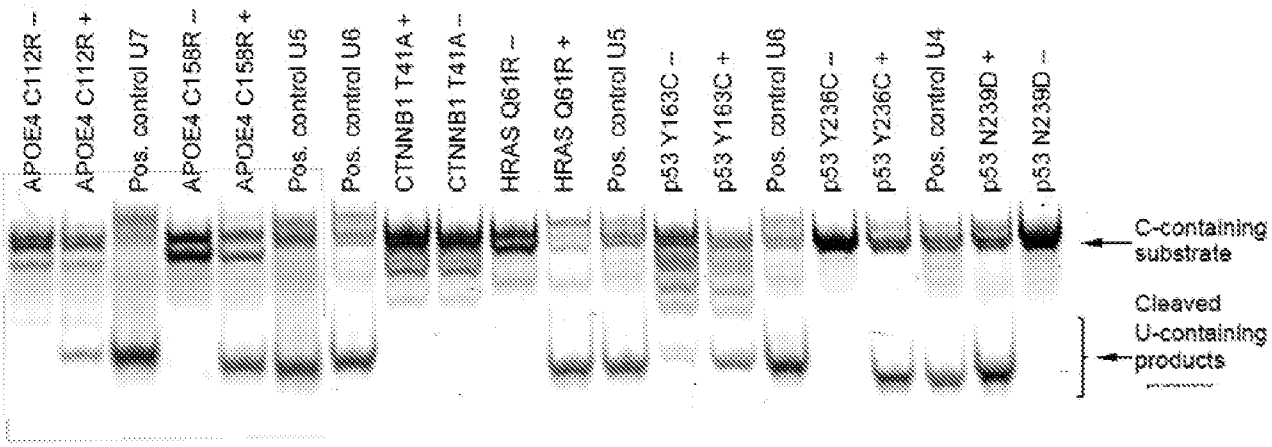


FIGURE 16B

Protospacer and PAM sequence: 5'-TTCCCCCCCCGATTTATTTATGG-3'

Sequence	% of total reads
CCCCCCCC	62.4
TTTTTTCC	18.2
TTTTTTTC	13.4
TTTTTTTT	3.3
TCCCCCCC	0.8
CCCCTTCC	0.3
CCCTTTCC	0.3
TTTTTCCC	0.3
CCCCTCCC	0.3

FIGURE 17

EMX1:	GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG
FANCF:	GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG
HEK293 site 2:	GAAC$_4$AC$_6$AAAGCATAGACTGCGGG
HEK293 site 3:	GGCC$_4$C$_5$AGACTGAGCACGTGATGG
HEK293 site 4:	GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2:	GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

FIGURE 18A

EMX1	C_5	C_6
NBE1	6.2%	6.5%
NBE1 + UGI	9.7%	10.1%
NBE2	8.0%	8.7%

FIGURE 18B

FANCF	C_6	C_7	C_8	C_{10}
NBE1	3.7%	3.2%	3.4%	2.4%
NBE1 + UGI	7.5%	7.6%	7.5%	1.6%
NBE2	4.7%	4.6%	4.6%	0.8%

FIGURE 18C

HEK293 site 2	C_4	C_6
NBE1	0.4%	0.4%
NBE1 + UGI	1.6%	2.6%
NBE2	3.4%	5.9%

FIGURE 18D

HEK293 site 3	C_4	C_5
NBE1	2.0%	1.9%
NBE1 + UGI	6.5%	6.7%
NBE2	10.0%	12.5%

FIGURE 18E

HEK293 site 4	C_5
NBE1	1.4%
NBE1 + UGI	5.4%
NBE2	8.2%

FIGURE 18F

RNF2	C_3	C_6
NBE1	0.7%	1.4%
NBE1 + UGI	3.4%	3.9%
NBE2	2.5%	3.7%

FIGURE 18G

Non-protospacer Cs	C	T
untreated	99.93%	0.03%
NBE1	99.95%	0.03%
NBE1 + UGI	99.91%	0.06%
NBE2	99.92%	0.04%

FIGURE 18H

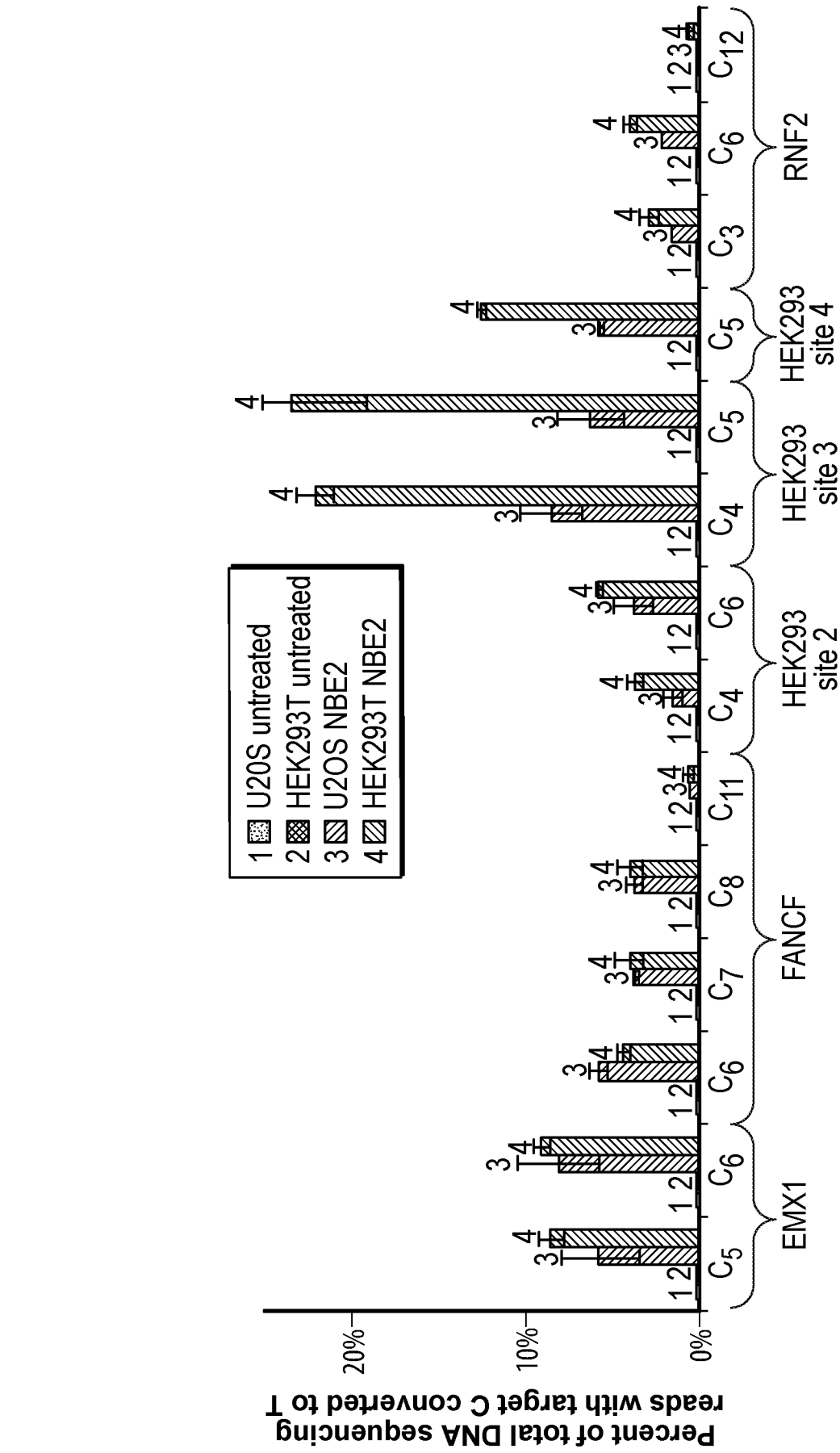

Percent of total DNA sequencing reads with target C converted to T
20%
10%
0%
1 U20S untreated
2 HEK293T untreated
3 U2OS NBE2
4 HEK293T NBE2
EMX1
FANCF
HEK293 site 2
HEK293 site 3
HEK293 site 4
RNF2

FIGURE 19

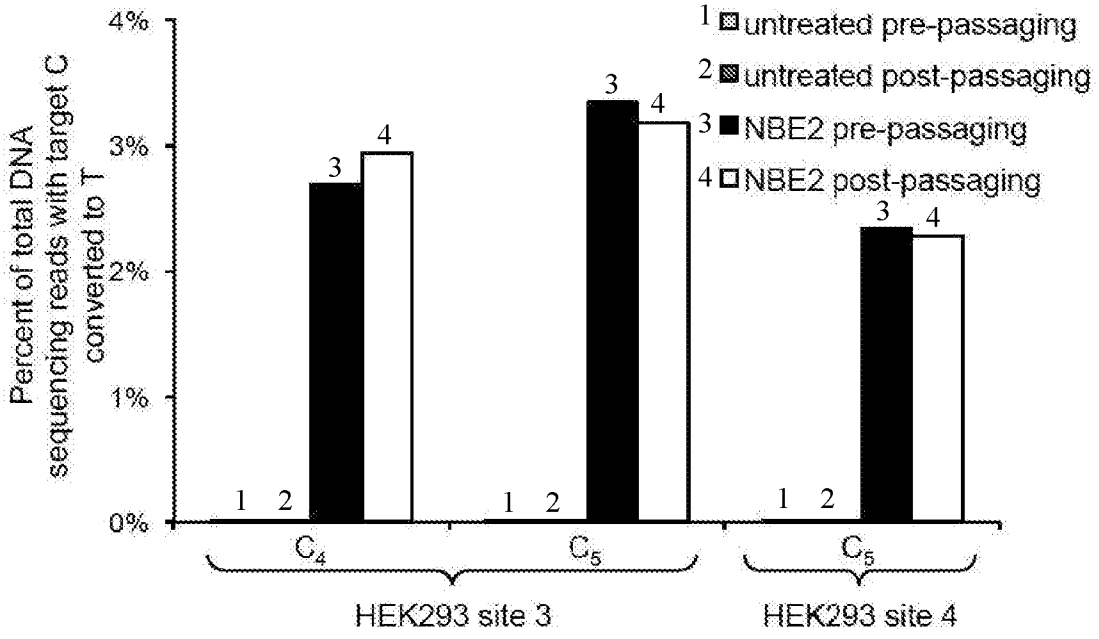


FIGURE 20

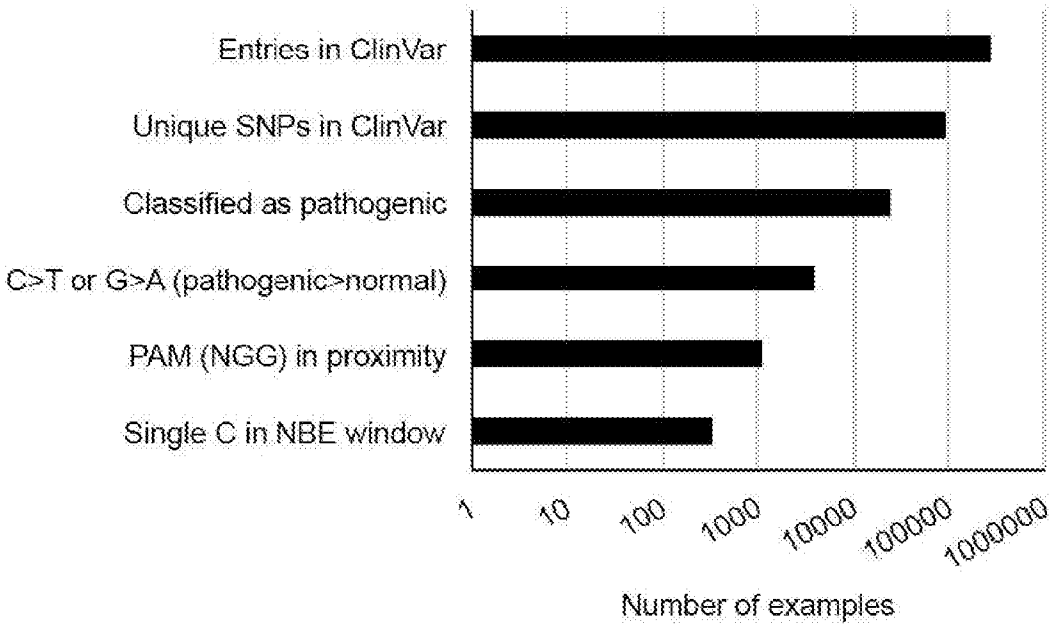


FIGURE 21

EMX1		G	A	G	T	C5	C6	G	A	G	C	A	G	A	A	G	A	A	G	A	A	G	G	G
1.85 uM NBE1	A	0.2	99.9	0.1	0.1	0.1	0.0	0.3	100.0	0.1	0.1	99.9	0.2	99.9	99.7	0.3	100.0	99.9	0.2	99.9	97.9	0.1	0.2	0.2
	C	0.0	0.0	0.0	0.0	60.4	61.0	0.0	0.0	0.0	99.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0
	G	99.7	0.0	99.9	0.0	0.0	0.0	99.7	0.0	99.9	0.1	0.0	99.7	0.0	0.0	99.7	0.0	0.0	99.8	0.0	0.0	99.8	99.8	99.8
	T	0.1	0.0	0.0	99.9	39.5	39.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

FANCF		G	G	A	A	T	C6	C7	C8	T	T	C11	T	G	C	A	G	C	A	C	C	T	G	G
1.85 uM NBE1	A	0.1	0.2	99.8	99.9	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	99.9	0.2	0.2	95.0	0.1	0.1	0.1	0.3	0.4
	C	0.0	0.0	0.1	0.1	0.0	63.9	64.7	65.0	0.0	0.0	72.6	0.0	0.0	99.4	0.1	0.0	99.7	4.9	99.8	99.9	0.0	0.0	0.0
	G	99.8	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.2	0.0	99.7	0.2	0.0	0.1	0.0	0.0	99.6	99.5
	T	0.1	0.1	0.0	0.0	99.9	36.0	35.1	34.9	99.9	99.9	27.3	99.8	0.1	0.3	0.0	0.1	0.0	0.0	0.0	0.0	99.9	0.1	0.1

HEK293 site 2		G	A	A	C4	A	C6	A	A	A	G	C	A	T	A	G	A	C	T	G	C	G	G	G
1.85 uM NBE1	A	0.1	100.0	99.8	0.1	100.0	0.1	99.9	100.0	99.9	0.1	0.1	100.0	0.0	99.9	0.2	100.0	0.1	0.0	0.1	0.1	0.1	0.2	0.2
	C	0.0	0.0	0.1	80.6	0.0	76.9	0.0	0.0	0.0	0.0	99.6	0.0	0.0	0.1	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0
	G	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.1	0.0	99.8	0.0	0.0	0.0	99.8	0.0	99.9	99.8	99.8
	T	0.1	0.0	0.0	19.3	0.0	22.9	0.0	0.0	0.0	0.1	0.3	0.0	99.9	0.0	0.1	0.0	0.0	99.9	0.0	0.1	0.0	0.1	0.0

HEK293 site 3		G	G	C	C4	C5	A	G	A	C	T	G	A	G	C	A	C	G	T	G	A	T	G	G
1.85 uM NBE1	A	0.2	0.1	0.1	0.1	0.0	99.9	0.2	99.9	0.1	0.0	0.2	99.9	0.1	0.1	100.0	0.1	0.1	0.0	0.2	98.6	0.0	0.1	0.1
	C	0.0	0.0	92.2	74.6	71.5	0.0	0.0	0.0	96.6	0.0	0.0	0.0	0.0	99.8	0.0	99.7	0.0	0.0	0.0	1.3	0.0	0.0	0.0
	G	99.7	99.8	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	99.8	0.0	99.8	0.0	0.0	0.0	99.8	0.0	99.8	0.0	0.0	99.9	99.8
	T	0.1	0.1	7.7	25.1	28.5	0.0	0.0	0.0	3.3	99.9	0.1	0.0	0.0	0.1	0.0	0.2	0.0	99.9	0.0	0.0	100.0	0.0	0.0

FIGURE 22

HEK293 site 3		G	G	C	C_4	C_5	A	G	A	C	T	G	A	G	C	A	C	G	T	G	A	T	G	G
1.85 uM NBE1	A	0.2	0.1	0.1	0.1	0.0	99.9	0.2	99.9	0.1	0.0	0.2	99.9	0.1	0.1	100.0	0.1	0.1	0.0	0.2	98.6	0.0	0.1	0.1
	C	0.0	0.0	92.2	74.8	71.5	0.0	0.0	0.0	96.6	0.0	0.0	0.0	0.0	99.8	0.0	99.7	0.0	0.0	0.0	1.3	0.0	0.0	0.0
	G	99.7	99.8	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	99.8	0.0	99.8	0.0	0.0	0.0	99.8	0.0	99.8	0.0	0.0	99.9	99.8
	T	0.1	0.1	7.7	25.1	28.5	0.0	0.0	0.0	3.3	99.9	0.1	0.0	0.0	0.1	0.0	0.2	0.0	99.9	0.0	0.0	100.0	0.0	0.0

HEK293 site 4		G	G	C	A	C_5	T	G	C	G	G	C	T	G	G	A	G	G	T	G	G	G	G	G
1.85 uM NBE1	A	0.3	0.2	0.1	100.0	0.1	0.0	0.2	0.1	0.2	0.2	0.1	0.0	0.3	0.5	100.0	0.1	0.1	0.0	0.1	0.1	0.1	0.2	0.2
	C	0.0	0.0	98.8	0.0	60.1	0.0	0.0	97.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.7	99.7	0.0	0.0	0.0	0.0	99.8	0.0	99.8	99.8	0.0	0.0	99.7	99.4	0.0	99.8	99.8	0.1	99.8	99.8	99.8	99.8	99.7
	T	0.0	0.1	1.1	0.0	39.8	100.0	0.1	2.9	0.0	0.1	0.5	100.0	0.0	0.1	0.0	0.0	0.0	99.9	0.0	0.1	0.1	0.0	0.0

RNF2		G	T	C_3	A	T	C_6	T	T	A	G	T	C_{12}	A	T	T	A	C	C	T	G	A	G	G
1.85 uM NBE1	A	0.1	0.0	0.1	99.9	0.0	0.0	0.0	0.0	99.9	0.1	0.0	0.1	99.9	0.0	0.0	99.9	0.1	0.0	0.0	0.1	98.6	0.0	0.0
	C	0.0	0.0	59.1	0.0	0.0	57.8	0.0	0.0	0.1	0.0	0.0	86.4	0.0	0.0	0.0	0.1	99.9	99.9	0.0	0.0	1.4	0.0	0.0
	G	99.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	99.9	99.9
	T	0.0	99.9	40.8	0.0	100.0	42.1	100.0	100.0	0.0	0.1	99.9	13.4	0.0	100.0	99.9	0.0	0.0	0.0	99.9	0.1	0.0	0.1	0.0

FIGURE 22 (CONTINUED)

																								GUIDE-seq counts
EMX1 on target	G	A	G	T	C	C	G	A	G	C	A	G	A	A	G	A	A	G	A	A	G	G	G	4521
untreated					0.1	0.0				0.0														
NBE1					4.2	4.4				0.0														
NBE2					9.1	9.3				0.0														
NBE3					26.0	26.6				0.0														
EMX1 off target 1	G	A	G	T	T	A	G	A	G	C	A	G	A	A	G	A	A	G	A	A	A	G	G	3123
untreated										0.0														
NBE1										0.0														
NBE2										0.0														
NBE3										0.0														
EMX1 off target 2	G	A	G	T	C	T	A	A	G	C	A	G	A	A	G	A	A	G	A	A	G	A	G	1445
untreated					0.0					0.0														
NBE1					1.0					0.0														
NBE2					3.6					0.0														
NBE3					8.5					0.0														

FIGURE 23

FIGURE 23 (CONTINUED)

EMX1 off target 7	G	A	G	C	C	G	G	A	G	C	C	A	G	A	A	G	G	A	G	G	143
untreated				0.0	0.0						0.0										
NBE1				0.0	0.0						0.0										
NBE2				0.1	0.0						0.0										
NBE3				0.1	0.1						0.0										

EMX1 off target 8	A	A	G	G	T	C	C	G	A	G	G	A	G	A	G	G	A	A	A	G	G	102
untreated					0.0	0.0																
NBE1					0.0	0.0																
NBE2					0.0	0.0																
NBE3					0.1	0.1																

EMX1 off target 9	G	A	A	T	C	C	A	A	G	C	A	G	G	A	G	A	A	A	G	G	A	67
untreated					0.0	0.1					0.0											
NBE1					0.0	0.1					0.0											
NBE2					0.0	0.1					0.0											
NBE3					0.3	0.3					0.0											

EMX1 off target 10	A	C	G	T	C	T	G	A	G	C	A	G	A	G	A	A	G	A	A	T	G	G	38
untreated		0.0			0.1						0.0												
NBE1		0.0			0.3						0.0												
NBE2		0.0			0.5						0.0												
NBE3		0.0			2.1						0.0												

FIGURE 23 (CONTINUED)

																						GUIDE-seq counts
FANCF on target	G	G	A	A	T	C	C	C	T	T	C	T	G	C	A	G	C	A	C	C	T	G	G	4616
untreated						0.0	0.0	0.0			0.0			0.0			0.0		0.0	0.0				
NBE1						3.7	3.2	3.4			2.4			0.0			0.0		0.0	0.0				
NBE2						4.1	3.5	3.5			0.4			0.0			0.0		0.0	0.0				
NBE3						19.1	16.0	16.7			0.8			0.1			0.0		0.0	0.0				
FANCF off target 1	G	G	A	A	C	C	C	C	G	T	C	T	G	C	A	G	C	A	C	C	A	G	G	2899
untreated					0.0	0.0	0.0	0.0			0.0			0.0			0.0		0.0	0.0				
NBE1					0.0	0.4	0.0	0.1			0.0			0.0			0.0		0.0	0.0				
NBE2					0.2	0.2	0.3	0.3			0.0			0.1			0.0		0.0	0.0				
NBE3					1.9	2.1	1.5	1.3			1.3			0.0			0.0		0.0	0.0				
FANCF off target 2	G	G	A	G	T	C	C	C	T	C	C	T	A	C	A	G	C	A	C	C	A	G	G	524
untreated						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0				
NBE1						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0				
NBE2						0.1	0.1	0.1		0.1	0.1			0.0			0.0		0.0	0.0				
NBE3						0.2	0.2	0.1		0.1	0.1			0.0			0.0		0.0	0.0				
FANCF off target 3	A	G	A	G	G	C	C	C	C	T	C	T	G	C	A	G	C	A	C	C	A	G	G	150
untreated						0.0	0.1	0.1	0.0		0.0			0.0			0.0		0.0	0.0				
NBE1						0.0	0.1	0.1	0.0		0.0			0.0			0.0		0.0	0.0				
NBE2						0.0	0.1	0.1	0.0		0.0			0.0			0.1		0.0	0.0				
NBE3						0.0	0.0	0.0	0.0		0.0			0.0			0.0		0.0	0.0				
FANCF off target 4	A	C	C	A	T	C	C	C	T	C	C	T	G	C	A	G	C	A	C	C	A	G	G	125
untreated		0.0	0.0			0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0				
NBE1		0.0	0.0			0.0	0.0	0.1		0.1	0.0			0.0			0.0		0.0	0.2				
NBE2		0.0	0.0			0.1	0.1	0.1		0.2	0.0			0.0			0.0		0.0	0.0				
NBE3		0.0	0.0			0.7	0.4	0.3		0.1	0.0			0.0			0.0		0.0	0.0				

FIGURE 24

FANCF off target 5	G	G	A	T	T	G	C	C	A	T	C	C	G	C	A	G	C	A	C	C	T	G	G	101
untreated							0.1	0.0			0.0	0.0		0.0			0.0		0.0	0.0				
NBE1							0.1	0.0			0.0	0.1		0.0			0.0		0.0	0.0				
NBE2							0.1	0.0			0.0	0.0		0.0			0.0		0.0	0.0				
NBE3							0.1	0.0			0.0	0.0		0.0			0.0		0.0	0.0				
FANCF off target 6	T	G	A	A	T	C	C	C	A	T	C	T	C	C	A	G	C	A	C	C	A	G	G	78
untreated						0.0	0.0	0.0			0.0		0.1	0.1			0.0		0.0	0.0				
NBE1						0.0	0.0	0.0			0.0		0.0	0.0			0.0		0.0	0.0				
NBE2						0.2	0.2	0.2			0.2		0.1	0.0			0.0		0.0	0.0				
NBE3						0.3	0.3	0.3			0.3		0.1	0.0			0.0		0.0	0.0				
FANCF off target 7	G	G	A	G	T	C	C	C	T	C	C	T	A	C	A	G	C	A	C	C	A	G	G	77
untreated						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0				
NBE1						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0				
NBE2						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0				
NBE3						0.5	0.5	0.5		0.3	0.1			0.0			0.0		0.0	0.0				
FANCF off target 8	G	G	A	G	T	C	C	C	T	C	C	T	G	C	A	G	C	A	C	C	T	G	A	4
untreated						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0				
NBE1						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0				
NBE2						1.4	0.5	0.5		0.4	0.5			0.0			0.0		0.0	0.0				
NBE3						0.1	0.1	0.0		0.0	0.0			0.0			0.0		0.0	0.0				

FIGURE 24 (CONTINUED)

																						GUIDE-seq counts
HEK site 2 on target	G	A	A	C	A	C	A	A	A	G	C	A	T	A	G	A	C	T	G	C	G	G	G	1670
untreated				0.0		0.0					0.0						0.0			0.0				
NBE1				1.8		2.9					0.0						0.0			0.0				
NBE2				9.5		10.5					0.0						0.0			0.0				
NBE3				17.6		14.9					0.0						0.0			0.1				
HEK site 2 off target 1	G	A	A	C	A	C	A	A	T	G	C	A	T	A	G	A	T	T	G	C	C	G	G	99
untreated				0.0		0.0					0.0									0.1	0.0			
NBE1				0.0		0.0					0.1									0.1	0.2			
NBE2				0.0		0.4					0.0									0.1	0.0			
NBE3				0.2		0.6					0.0									0.1	0.0			
HEK site 2 off target 2	A	A	A	C	A	T	A	A	A	G	C	A	T	A	G	A	C	T	G	C	A	A	A	4
untreated				0.0							0.0						0.0			0.0				
NBE1				0.0							0.0						0.0			0.0				
NBE2				0.0							0.0						0.0			0.0				
NBE3				0.0							0.0						0.0			0.0				

FIGURE 25

																						GUIDE-seq counts
HEK site 3 on target	G	G	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	A	T	G	G	2074
untreated			0.1	0.1	0.1				0.0					0.0		0.0								
NBE1			0.1	4.8	4.3				0.0					0.0		0.0								
NBE2			3.6	21.5	20.4				0.0					0.0		0.0								
NBE3			0.5	35.4	33.1				0.5					0.0		0.2								
HEK site 3 off target 1	C	A	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	C	T	G	G	327
untreated	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0				
NBE1	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0				
NBE2	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0				
NBE3	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0				
HEK site 3 off target 2	G	A	C	A	C	A	G	A	C	T	G	G	G	C	A	C	G	T	G	A	G	G	G	306
untreated			0.0		0.0				0.0					0.0		0.0								
NBE1			0.0		0.0				0.0					0.0		0.0								
NBE2			0.0		0.0				0.0					0.0		0.0								
NBE3			0.0		0.0				0.0					0.0		0.0								
HEK site 3 off target 3	A	G	C	T	C	A	G	A	C	T	G	A	G	C	A	A	G	T	G	A	G	G	G	136
untreated			0.0		0.0				0.0					0.0										
NBE1			0.0		0.0				0.0					0.0										
NBE2			0.0		0.0				0.0					0.0										
NBE3			0.0		0.9				0.0					0.0										
HEK site 3 off target 4	A	G	A	C	C	A	G	A	C	T	G	A	G	C	A	A	G	A	G	A	G	G	G	10
untreated				0.0	0.0				0.0					0.0										
NBE1				0.0	0.0				0.0					0.1										
NBE2				0.0	0.0				0.0					0.0										
NBE3				0.0	0.0				0.0					0.0										
HEK site 3 off target 5	G	A	G	C	C	A	G	A	A	T	G	A	G	C	A	C	G	T	G	A	G	G	G	2
untreated				0.0	0.0									0.0		0.0								
NBE1				0.0	0.0									0.0		0.0								
NBE2				0.0	0.0									0.0		0.0								
NBE3				0.0	0.0									0.0		0.0								

FIGURE 26

																						GUIDE-seq counts
HEK site 4 on target	G	G	C	A	C	T	G	C	G	G	C	T	G	G	A	G	G	T	G	G	G	G	1054
untreated			0.0		0.0			0.1			0.0												
NBE1			0.0		0.4			0.1			0.0												
NBE2			0.1		12.7			0.1			0.0												
NBE3			1.5		22.5			1.0			0.0												
HEK site 4 off target 1	T	G	C	A	C	T	G	C	G	G	C	C	G	G	A	G	G	A	G	G	T	G	G	2475
untreated			0.0		0.0			0.0			0.0	0.0											
NBE1			0.0		0.6			0.1			0.1	0.0											
NBE2			0.3		3.3			0.3			0.0	0.0											
NBE3			0.2		10.2			0.3			0.0	0.0											
HEK site 4 off target 2	G	G	C	T	C	T	G	C	G	G	C	T	G	G	A	G	G	G	G	G	T	G	G	2020
untreated			0.0		0.0			0.1			0.0												
NBE1			0.0		0.6			0.1			0.0												
NBE2			0.0		0.4			0.1			0.1												
NBE3			0.0		3.3			0.1			0.0												
HEK site 4 off target 3	G	G	C	A	C	G	A	C	G	G	C	T	G	G	A	G	G	T	G	G	G	G	G	1375
untreated			0.0		0.0			0.0			0.0												
NBE1			0.1		11.8			4.0			0.0												
NBE2			0.0		9.6			3.9			0.0												
NBE3			0.1		19.0			1.9			0.1												
HEK site 4 off target 4	G	G	C	A	T	C	A	C	G	G	C	T	G	G	A	G	G	T	G	G	A	G	G	1097
untreated			0.0			0.0		0.0			0.0												
NBE1			0.0			0.2		0.3			0.0												
NBE2			0.5			5.3		2.7			0.0												
NBE3			0.2			16.1		3.1			0.0												

FIGURE 27

HEK site 4 off target 5	G	G	C	G	C	T	G	C	G	G	C	G	G	G	A	G	G	T	G	G	A	G	G	1002
untreated			0.0		0.0			0.1			0.0													
NBE1			0.0		0.0			0.0			0.0													
NBE2			0.0		0.0			0.1			0.0													
NBE3			0.1		0.1			0.1			0.0													
HEK site 4 off target 6	G	G	C	A	C	T	G	A	G	A	C	T	G	G	G	G	G	T	G	G	G	G	G	982
untreated			0.0		0.0						0.0													
NBE1			0.0		0.5						0.3													
NBE2			0.8		3.6						0.0													
NBE3			0.0		6.6						0.0													
HEK site 4 off target 7	A	G	C	A	G	T	G	C	G	G	C	T	A	G	A	G	G	T	G	G	T	G	G	981
untreated			0.0					0.1			0.0													
NBE1			0.0					0.1			0.0													
NBE2			0.0					0.1			0.0													
NBE3			0.0					0.2			0.0													
HEK site 4 off target 8	G	G	C	A	C	T	G	C	T	A	C	T	G	G	G	G	G	T	G	G	T	G	G	789
untreated			0.0		0.0			0.0			0.0													
NBE1			0.0		2.1			0.0			0.0													
NBE2			0.2		5.1			0.5			0.0													
NBE3			0.2		6.2			0.1			0.0													
HEK site 4 off target 9	G	G	C	A	C	T	G	G	G	G	C	T	G	G	G	G	G	A	G	G	G	G	G	712
untreated			0.0		0.0						0.0													
NBE1			0.0		0.3						0.0													
NBE2			0.0		0.4						0.0													
NBE3			0.0		0.9						0.0													
HEK site 4 off target 10	G	G	C	A	C	T	G	G	G	G	T	T	G	G	A	G	G	T	G	G	G	G	G	664
untreated			0.0		0.0																			
NBE1			0.0		0.0																			
NBE2			0.6		3.5																			
NBE3			0.3		10.0																			

FIGURE 27 (CONTINUED)

Non-protospacer Cs	C (%)	T (%)
untreated	99.94	0.04
NBE1	99.92	0.05
NBE2	99.92	0.05
NBE3	99.94	0.03

FIGURE 28

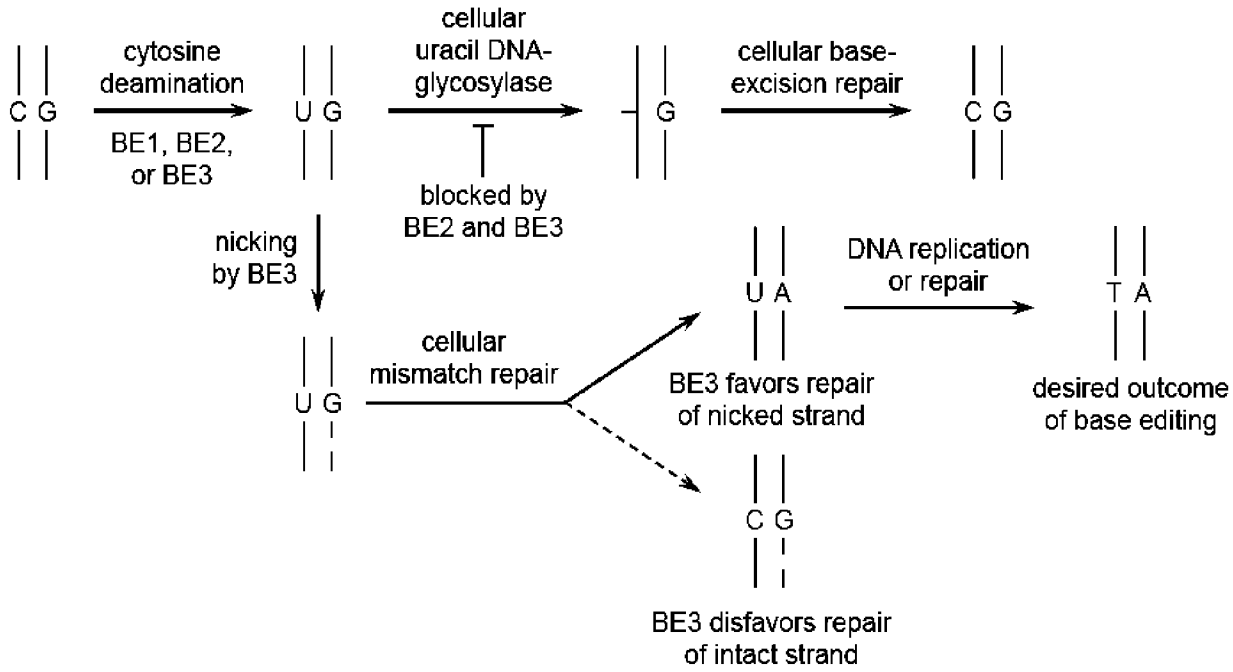


FIGURE 29A

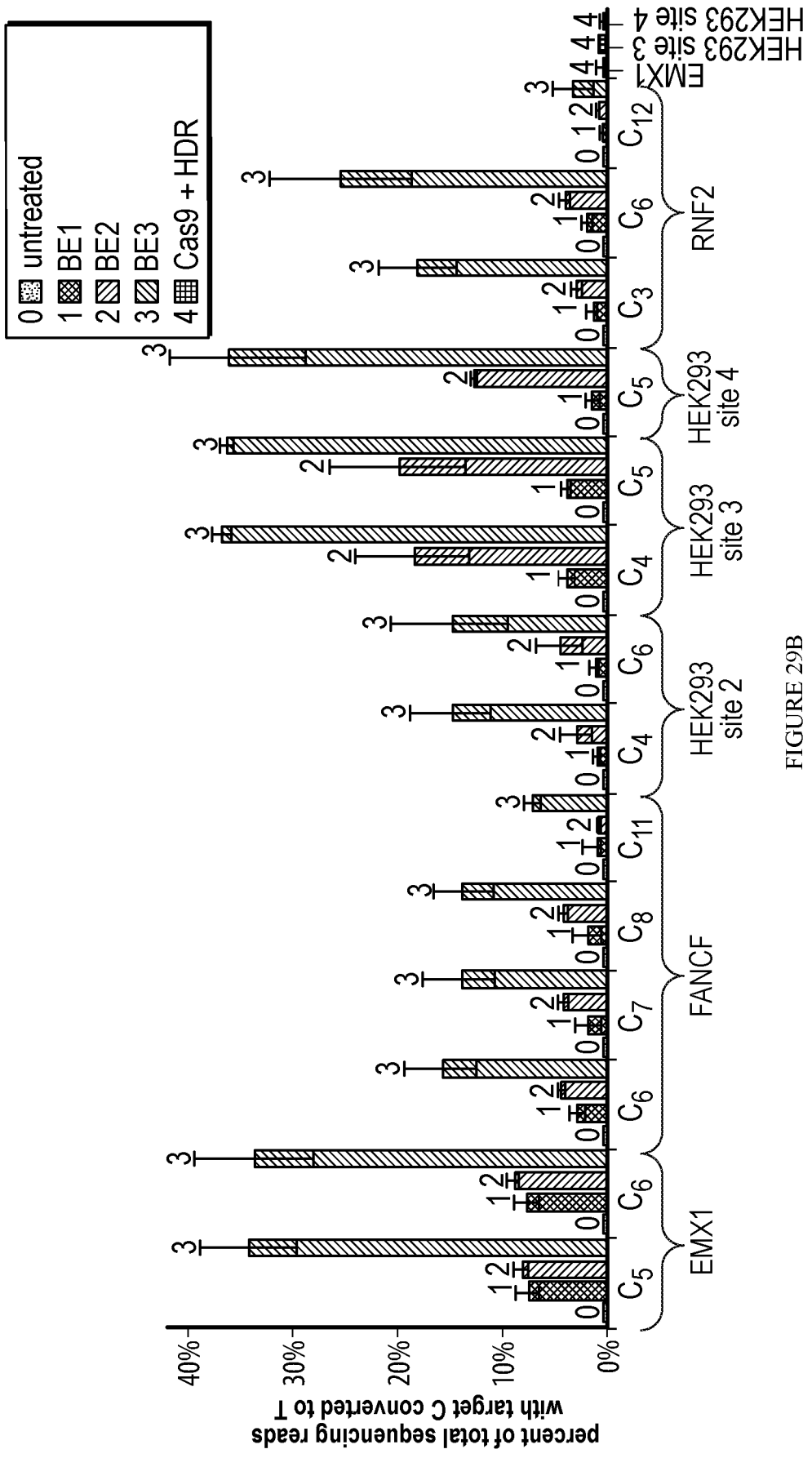

0 untreated
1 BE1
2 BE2
3 BE3
4 Cas9 + HDR
percent of total sequencing reads with target C converted to T
40%
30%
20%
10%
0%
EMX1
FANCF
HEK293 site 2
HEK293 site 3
HEK293 site 4
RNF2
EMX1
HEK293 site 3
HEK293 site 4

FIGURE 29B

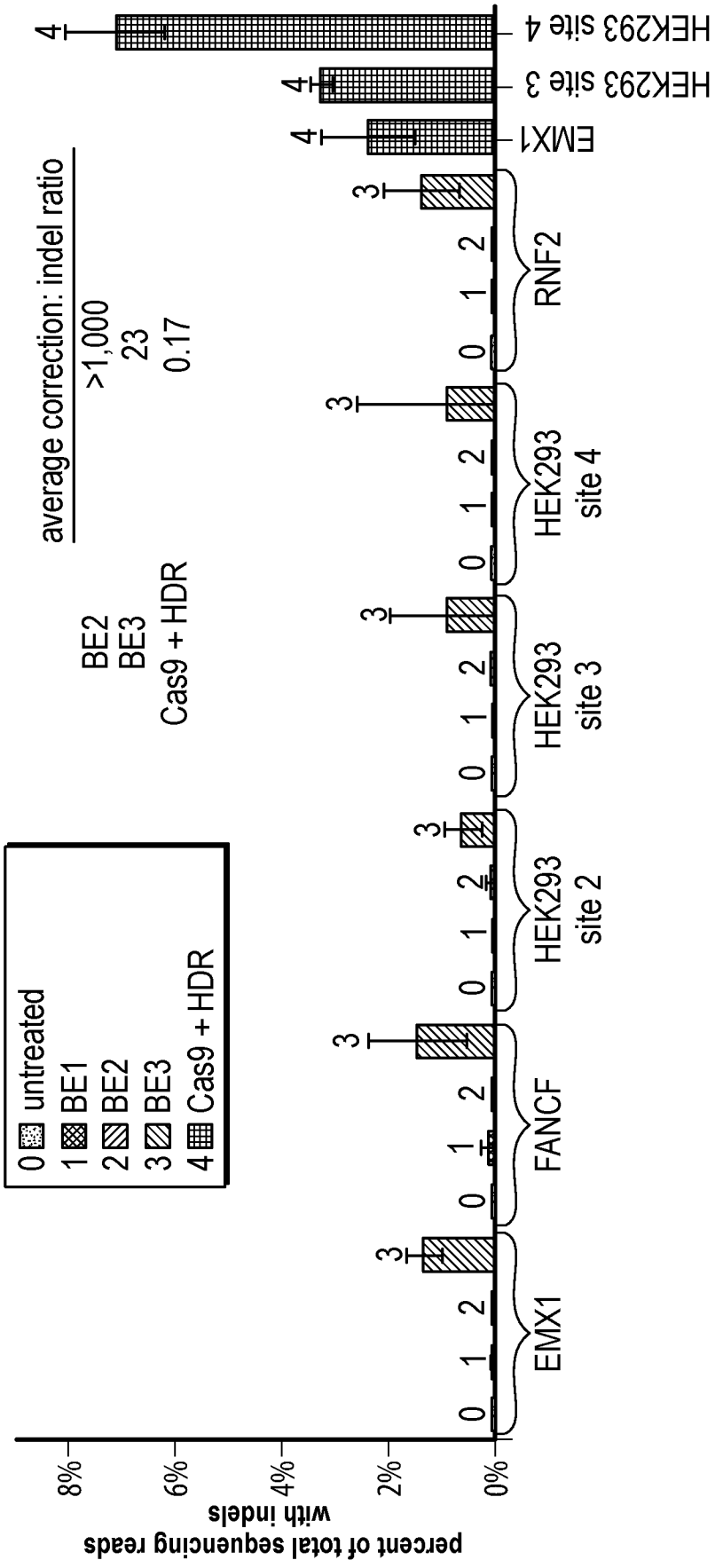

percent of total sequencing reads with indels
8%
6%
4%
2%
0%
0 untreated
1 BE1
2 BE2
3 BE3
4 Cas9 + HDR
average correction: indel ratio
BE2 >1,000
BE3 23
Cas9 + HDR 0.17
EMX1
FANCF
HEK293 site 2
HEK293 site 3
HEK293 site 4
RNF2
EMX1
HEK293 site 3
HEK293 site 4

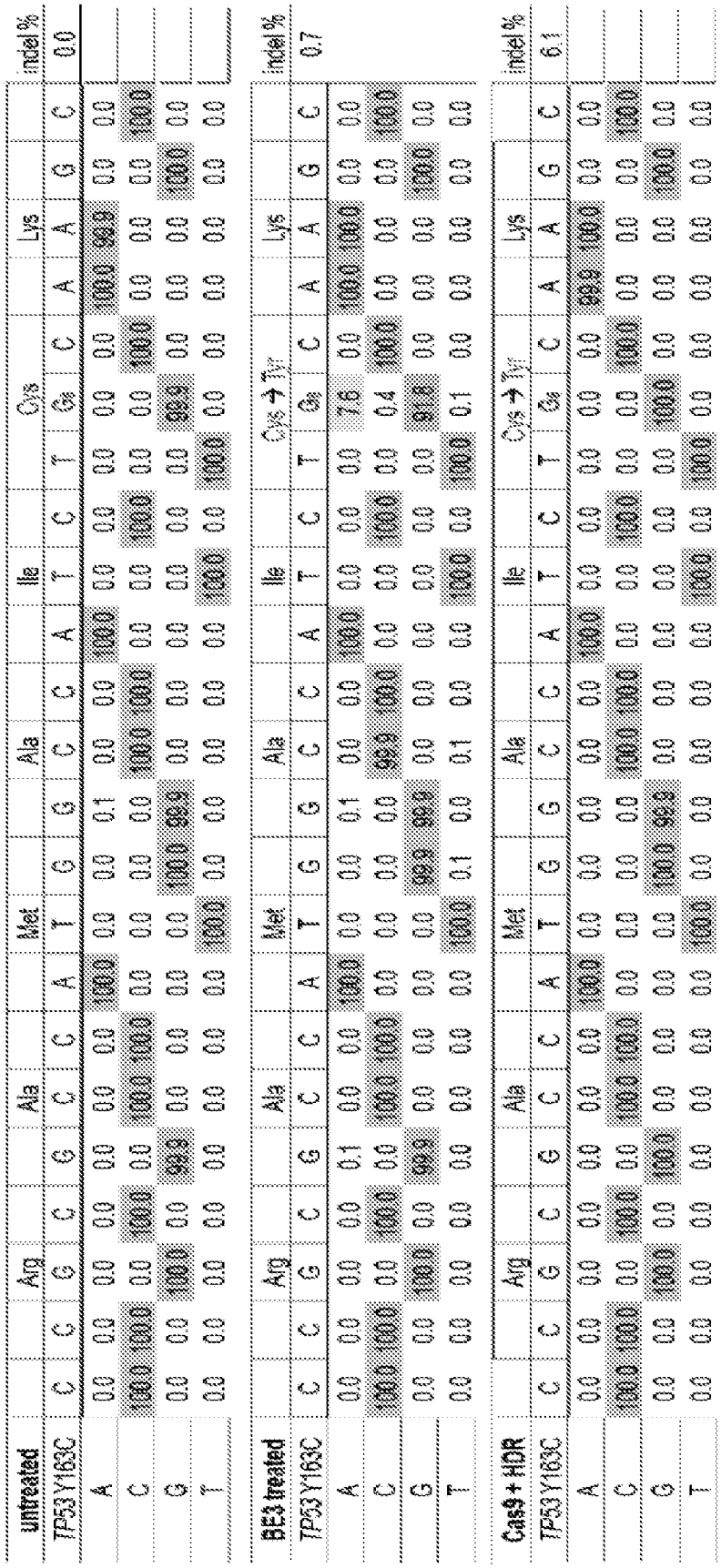
FIGURE 29C

untreated		Lys			Arg			Leu			Ala			Val			Tyr				Gln			indel %
APOE4 C158R	G	A	A	G	C_5	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G	0.0
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	
C	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

BE3 treated		Lys			Arg → Cys			Leu → Leu			Ala			Val			Tyr				Gln			indel %
APOE4 C158R	G	A	A	G	C_5	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G	4.6
A	0.1	100.0	100.0	0.0	0.5	0.0	1.3	0.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.1	
C	0.0	0.0	0.0	0.0	23.7	0.0	47.4	43.5	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
G	99.9	0.0	0.0	100.0	0.9	99.9	1.1	0.7	0.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	99.9	
T	0.0	0.0	0.0	0.0	74.9	0.1	50.2	55.0	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

Cas9 + HDR		Lys			Arg → Cys			Leu			Ala			Val			Tyr				Gln			indel %
APOE4 C158R	G	A	A	G	C_5	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G	26.1
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	
C	0.0	0.0	0.0	0.0	99.7	0.0	99.9	99.9	0.0	0.0	0.0	100.0	0.5	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	99.6	0.6	99.9	0.2	0.0	0.0	0.0	0.0	100.0	100.0	
T	0.0	0.0	0.0	0.0	0.3	0.0	0.1	0.1	100.0	0.0	0.0	0.0	0.1	0.4	99.3	0.1	99.8	0.0	0.0	0.0	0.0	0.0	0.0	

FIGURE 30A



untreated			Arg				Ala				Met				Ala				Ile				Cys				Lys				indel %
TP53Y163C	C	C	G	C	G	C	C	A	T	G	G	C	A	T	C	T	G$_e$	C	A	A	G	C								
A	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0								
C	100.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0								
G	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	0.0								
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0								

BE3 treated			Arg				Ala				Met				Ala				Ile				Cys → Tyr				Lys				indel %
TP53Y163C	C	C	G	C	G	C	C	A	T	G	G	C	A	T	C	T	G$_e$	C	A	A	G	C								
A	0.0	0.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	7.6	0.0	100.0	100.0	0.0	0.0	0.0				0.7
C	100.0	100.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	99.9	100.0	0.0	0.0	0.0	100.0	0.1	0.0	0.4	100.0	0.0	0.0	0.0	100.0					
G	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0	91.8	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0					
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	0.1	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0					

Cas9 + HDR			Arg				Ala				Met				Ala				Ile				Cys → Tyr				Lys				indel %
TP53Y163C	C	C	G	C	G	C	C	A	T	G	G	C	A	T	C	T	G$_e$	C	A	A	G	C								
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	0.0								6.1
C	100.0	100.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0								
G	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0								
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0								

FIGURE 30B

																							GUIDE-seq counts/ ChIP-seq intensity
HEK site 2 on target	G	A	A	C	A	C	A	A	A	G	C	A	T	A	G	A	C	T	G	C	G	G	G	1670/ 117
untreated				0.0		0.0					0.0						0.0			0.0				
BE1				1.8		2.9					0.0						0.0			0.0				
BE2				8.5		10.5					0.0						0.0			0.0				
BE3				17.6		14.9					0.0						0.0			0.1				
HEK site 2 GUIDE-seq off target 1	G	A	A	C	A	C	A	A	T	G	C	A	T	A	G	A	T	T	G	C	C	G	G	99
untreated				0.0		0.0					0.0									0.1	0.0			
BE1				0.0		0.0					0.1									0.1	0.2			
BE2				0.0		0.4					0.0									0.1	0.0			
BE3				0.2		0.6					0.0									0.1	0.0			
HEK site 2 GUIDE-seq off target 2	A	A	A	C	A	T	A	A	A	G	C	A	T	A	G	A	C	T	G	C	A	A	A	4
untreated				0.0							0.0						0.0			0.0				
BE1				0.0							0.0						0.0			0.0				
BE2				0.0							0.0						0.0			0.0				
BE3				0.0							0.0						0.0			0.0				
HEK site 2 ChIP-seq off target 1	T	C	A	G	G	G	T	G	A	G	C	A	T	A	G	A	C	T	G	C	C	G	G	31
untreated		0.0									0.0						0.0			0.0	0.0			
BE1		0.0									0.0						0.0			0.1	0.0			
BE2		0.0									0.0						0.0			0.1	0.0			
BE3		0.0									0.0						0.0			0.1	0.0			

FIGURE 31

HEK site 2 ChIP-seq off target 2	G	A	A	T	C	G	G	G	G	A	G	G	G	G	A	G	G	A	C	T	G	C	G	G	G	28
untreated					0.0														0.0			0.0				
BE1					0.0														0.0			0.1				
BE2					0.0														0.0			0.1				
BE3					0.0														0.0			0.1				
HEK site 2 ChIP-seq off target 3	T	G	A	A	G	T	G	T	T	G	C	A	T	A	G	A	C	T	G	C	A	G	G			23
untreated											0.0						0.0			0.0						
BE1											0.0						0.0			0.1						
BE2											0.0						0.0			0.0						
BE3											0.0						0.0			0.0						
HEK site 2 ChIP-seq off target 4	G	G	A	G	A	G	A	G	A	G	C	A	T	A	G	A	C	T	G	C	T	G	G			16
untreated											0.0						0.0			0.0						
BE1											0.0						0.0			0.0						
BE2											0.0						0.0			0.0						
BE3											0.0						0.0			0.0						
HEK site 2 ChIP-seq off target 5	C	C	A	A	A	C	A	A	A	A	C	A	T	A	G	A	C	T	G	C	T	G	G			10
untreated	0.0	0.0				0.0					0.0						0.0			0.0						
BE1	0.0	0.0				0.0					0.0						0.0			0.0						
BE2	0.0	0.0				0.0					0.0						0.0			0.0						
BE3	0.0	0.0				0.0					0.0						0.0			0.2						

FIGURE 31 (CONTINUED)

HEK site 3 on target	G	G	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	A	T	G	G
untreated			0.1	0.1	0.1				0.0					0.0		0.0							
BE1			0.1	4.8	4.3				0.0					0.0		0.0							
BE2			3.6	21.3	20.4				0.0					0.0		0.0							
BE3			0.5	39.4	38.1				0.5					0.0		0.2							
GUIDE-seq counts/ ChIP-seq intensity	2074/ 163																						

HEK site 3 GUIDE-seq off target 1	C	A	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	C	T	G	G	327
untreated	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0				
BE1	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0				
BE2	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0				
BE3	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0				

HEK site 3 GUIDE-seq off target 2	G	A	C	A	C	A	G	A	C	T	G	G	G	C	A	C	G	T	G	A	G	G	G	306
untreated			0.0		0.0				0.0					0.0		0.0								
BE1			0.0		0.0				0.0					0.0		0.0								
BE2			0.0		0.0				0.0					0.0		0.0								
BE3			0.0		0.0				0.0					0.0		0.0								

HEK site 3 GUIDE-seq off target 3	A	G	C	T	C	A	G	A	C	T	G	A	G	C	A	A	G	T	G	A	G	G	G	135
untreated			0.0		0.0				0.0					0.0										
BE1			0.0		0.0				0.0					0.0										
BE2			0.0		0.0				0.0					0.0										
BE3			0.0		0.0				0.0					0.0										

HEK site 3 GUIDE-seq off target 4	A	G	A	C	C	A	G	A	C	T	G	A	G	C	A	A	G	A	G	A	G	G	G	10
untreated				0.0	0.0				0.0					0.0										
BE1				0.0	0.0				0.0					0.1										
BE2				0.0	0.0				0.0					0.0										
BE3				0.0	0.0				0.0					0.0										

FIGURE 32

HEK site 3 GUIDE-seq off target 5	G	A	G	C	C	A	G	A	A	T	G	A	G	C	A	C	G	T	G	A	G	G	G	2
untreated				0.0	0.0											0.0								
BE1				0.0	0.0											0.0								
BE2				0.0	0.0											0.0								
BE3				0.0	0.0											0.0								
HEK site 3 ChIP-seq off target 1	C	A	G	G	A	A	G	C	T	G	G	A	G	C	A	C	G	T	G	A	G	G	G	47
untreated	0.0							0.0						0.0		0.0								
BE1	0.0							0.0						0.0		0.0								
BE2	0.0							0.0						0.0		0.0								
BE3	0.0							0.0						0.0		0.0								
HEK site 3 ChIP-seq off target 2	A	A	G	G	C	T	G	A	G	G	G	A	G	C	A	C	G	T	G	A	A	G	G	45
untreated					0.0									0.0		0.0								
BE1					0.0									0.0		0.0								
BE2					0.0									0.0		0.0								
BE3					0.0									0.0		0.0								
HEK site 3 ChIP-seq off target 3	G	T	C	A	G	G	G	G	A	A	G	A	G	C	A	C	G	T	G	A	C	G	G	33
untreated		0.0												0.0		0.0					0.0			
BE1		0.0												0.0		0.0					0.0			
BE2		0.0												0.0		0.0					0.0			
BE3		0.0												0.0		0.0					0.0			
HEK site 3 ChIP-seq off target 4	G	T	T	G	T	G	A	A	C	T	G	A	G	C	A	C	G	T	G	A	G	G	G	31
untreated									0.0					0.0		1.7								
BE1									0.0					0.0		0.7								
BE2									0.0					0.0		0.4								
BE3									0.0					0.0		0.4								
HEK site 3 ChIP-seq off target 5	A	T	A	T	T	T	G	C	T	G	G	A	G	C	A	C	G	T	G	A	A	G	G	31
untreated								0.0						0.0		0.0								
BE1								0.0						0.0		0.0								
BE2								0.0						0.0		0.0								
BE3								0.0						0.0		0.0								

FIGURE 32 (CONTINUED)

																							GUIDE-seq counts/ ChIP-seq intensity
HEK site 4 on target	G	G	C	A	C	T	G	C	G	G	C	T	G	G	A	G	G	T	G	G	G	G	1054/ 49
untreated			0.0		0.0			0.1			0.0												
BE1			0.0		0.4			0.1			0.0												
BE2			0.1		12.7			0.1			0.0												
BE3			1.5		22.3			1.0			0.0												
HEK site 4 GUIDE-seq off target 1	T	G	C	A	C	T	G	C	G	G	C	C	G	G	A	G	G	A	G	G	T	G	G	2475
untreated			0.0		0.0			0.0			0.0	0.0											
BE1			0.0		0.6			0.1			0.1	0.0											
BE2			0.3		3.3			0.3			0.0	0.0											
BE3			0.2		10.2			0.3			0.0	0.0											
HEK site 4 GUIDE-seq off target 2	G	G	C	T	C	T	G	C	G	G	C	T	G	G	A	G	G	G	G	G	T	G	G	2020
untreated			0.0		0.0			0.1			0.0												
BE1			0.0		0.6			0.1			0.0												
BE2			0.0		0.4			0.1			0.1												
BE3			0.0		3.3			0.1			0.0												
HEK site 4 GUIDE-seq off target 3	G	G	C	A	C	G	A	C	G	G	C	T	G	G	A	G	G	T	G	G	G	G	G	1375
untreated			0.0		0.0			0.0			0.0												
BE1			0.1		11.5			4.0			0.0												
BE2			0.0		9.6			3.9			0.0												
BE3			0.1		19.0			1.9			0.1												
HEK site 4 GUIDE-seq off target 4	G	G	C	A	T	C	A	C	G	G	C	T	G	G	A	G	G	T	G	G	A	G	G	1097
untreated			0.0			0.0		0.0			0.0												
BE1			0.0			0.2		0.3			0.0												
BE2			0.5			5.3		2.7			0.0												
BE3			0.2			18.1		3.1			0.0												

FIGURE 33

HEK site 4 GUIDE-seq off target 5	G	G	C	G	C	T	G	C	G	G	C	G	G	A	G	G	G	1002
untreated			0.0		0.0			0.1			0.0							
BE1			0.0		0.0			0.0			0.0							
BE2			0.0		0.0			0.1			0.0							
BE3			0.1		0.1			0.1			0.0							
HEK site 4 GUIDE-seq off target 6	G	G	C	A	C	T	G	A	G	A	C	T	G	G	G	G	G	982
untreated			0.0		0.0						0.0							
BE1			0.0		0.5						0.3							
BE2			0.0		3.6						0.0							
BE3			0.0		6.8						0.0							
HEK site 4 GUIDE-seq off target 7	A	G	C	A	G	T	G	C	G	G	C	T	A	G	A	G	T	981
untreated			0.0					0.1			0.0							
BE1			0.0					0.1			0.0							
BE2			0.0					0.1			0.0							
BE3			0.0					0.2			0.0							
HEK site 4 GUIDE-seq off target 8	G	G	C	A	C	T	G	C	T	A	C	T	G	G	G	T	G	769
untreated			0.0		0.0			0.0			0.0							
BE1			0.0		2.1			0.0			0.0							
BE2			0.2		5.1			0.5			0.0							
BE3			0.2		8.2			0.1			0.0							
HEK site 4 GUIDE-seq off target 9	G	G	C	A	C	T	G	G	G	G	C	T	G	G	G	A	G	712
untreated			0.0		0.0						0.0							
BE1			0.0		0.3						0.0							
BE2			0.0		0.4						0.0							
BE3			0.0		0.9						0.0							

FIGURE 33 (CONTINUED)

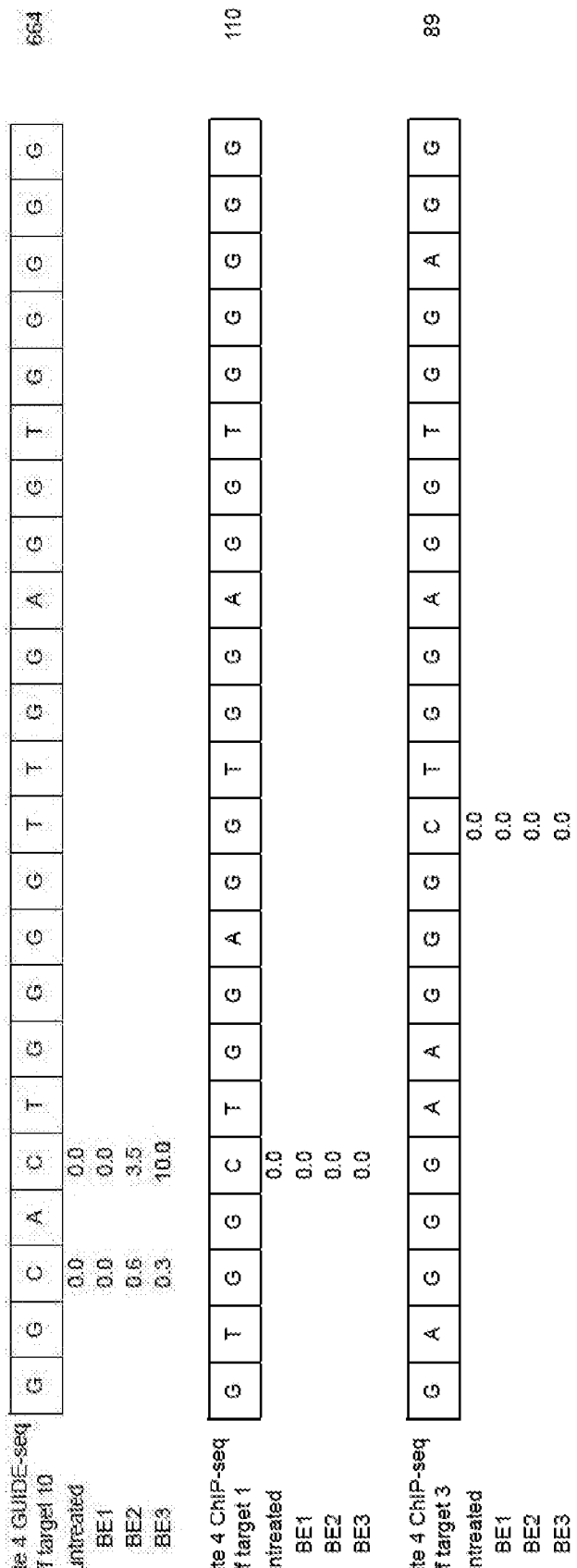


FIGURE 33 (CONTINUED)

APOE4	C50	A49	C48	C47	T46	G45	C44	G43	C42	A41	A40	G39	C38	T37	G36	C35	G34	T33	A32	A31	G30	C29	G28	G27	C26
Untreated																									
A	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	100±0.1
G	0.1±0.1	0.1±0.1	0±0	0±0	0.1±0.1	100±0.1	0±0	99.9±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0±0	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3 + on-target sgRNA																									
A	0±0	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1	0±0	0.1±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0±0	0±0	100±0.1
G	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3 + off-target sgRNA																									
A	0±0	100±0.1	0±0	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0±0	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0
C	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	100±0	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1
G	0±0	0.1±0.1	0±0	0±0	0±0	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0±0	100±0.1	0±0	100±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0±0	100±0.1	100±0.1	0±0
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0±0	0±0	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1

APOE4	T25	C24	C23	T22	C21	C20	G19	C18	G17	A16	T15	G14	C13	C12	G11	A10	T9	G8	A7	C6	C5	T4	G3	C2	A1
Untreated																									
A	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1
C	0.1±0.1	100±0	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0	100±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	100±0.1	0±0	0±0	0.1±0.1	0±0	0±0	100±0.1	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1
G	0±0	0±0	0±0	0.1±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0±0	0±0	100±0.1	0±0	0.1±0.1
T	100±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3 + on-target sgRNA																									
A	0±0	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0±0	0.1±0.1	0.2±0.1	100±0.1	0.1±0.1	0.2±0.1	99.9±0.2	0±0	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	99.9±0.2
C	0.1±0.1	99.9±0.1	100±0.1	0.1±0.1	99.5±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0±0	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.2±0.2
G	0±0	0±0	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	99.9±0.1	0.1±0.1	0±0	99.9±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1
T	100±0.1	0.2±0.1	0.1±0.1	100±0.1	0.5±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.9±0.2	0.1±0.1	0.1±0.1	0.1±0.1
BE3 + off-target sgRNA																									
A	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0	100±0.1	0±0	0.1±0.1	0.1±0.1	0±0	100±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
G	0±0	0±0	0±0	0.1±0.1	0±0	0.1±0.1	100±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1	0±0	0±0	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0±0	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1
T	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0±0

FIGURE 34

APOE4	A-1	G-2	G-3	C-4	C-5	G-6	G-7	G-8	G-9	C-10	C-11	C-12	G-13	C-14	G-15	A-16	G-17	G-18	G-19	C-20	G-21	C-22	C-23	G-24	A-25
Untreated																									
A	100±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.2±0.1	0.1±0.1	0±0	0.1±0.1	100±0.1
C	0.1±0.1	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	99.9±0.1	0.1±0.1	0.1±0.1	0±0	0±0	0±0	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0	0.1±0.1
G	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0±0	99.9±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1	99.9±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	99.9±0.1	0.1±0.1	0.1±0.1	100±0.2	0.1±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.2	0.1±0.1
BE3 + on-target sgRNA																									
A	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.2±0.1	0.2±0.2	0.2±0.1	0.2±0.1	0.1±0.1	0.1±0.1	99.9±0.1
C	0.1±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0±0	0±0	0.1±0.1	0±0	99.8±0.2	0±0	99.9±0.1	100±0.1	0±0	0±0
G	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.9±0.1	0±0	100±0.1	0.1±0.1	100±0.1	100±0.1	99.9±0.1	0.2±0.1	99.8±0.2	0±0	0±0	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.1
BE3 + off-target sgRNA																									
A	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	0.1±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0±0	0±0	0±0	100±0.1	100±0.1	100±0.1	0±0	100±0.1	0±0	0.1±0.1	0.1±0.1	0±0	0±0	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0	0±0
G	0.1±0.1	100±0.1	100±0.1	0±0	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0±0	0±0	0±0	99.9±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	99.9±0.1	0±0	0±0	100±0.1	0±0
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0±0	0.1±0.1	0±0	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1

APOE4	G-26	C-27	G-28	C-29	G-30	G-31	C-32	C-33	T-34	C-35	A-36	G-37	C-38	G-39	C-40	C-41	A-42	T-43	C-44	C-45	G-46	C-47	G-48	A-49	G-50
Untreated																									
A	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
C	0±0	100±0.1	0.1±0.1	99.9±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	99.6±0.1	0.1±0.1	100±0.1	100±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1
G	100±0.1	0.1±0.1	99.9±0.2	0.1±0.1	100±0.1	100±0.1	0±0	0±0	0±0	0±0	0.1±0.1	100±0.1	0.1±0.1	99.9±0.1	0±0	0±0	0.1±0.1	0.5±0.1	0.1±0.1	0.1±0.1	99.9±0.2	0±0	99.7±0.1	0.1±0.1	100±0.1
T	0.1±0.1	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.3±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.6±0.1	0.1±0.1	0.1±0.1	0.2±0.2	0.1±0.1	0.3±0.1	0±0	0.1±0.1
BE3 + on-target sgRNA																									
A	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
C	0±0	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	99.9±0.1	0±0	0±0	99.7±0.1	0.1±0.1	100±0.1	100±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0±0	0.2±0.2
G	100±0.1	0.1±0.1	100±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	99.8±0.2	0.1±0.1	0±0	0.1±0.1	0.5±0.5	0±0	0±0	100±0.1	0.1±0.1	99.6±0.4	0.1±0.1	99.9±0.2
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	0.2±0.1	0.1±0.1	0.1±0.1	0.4±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.6±0.5	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.4±0.3	0±0	0±0
BE3 + off-target sgRNA																									
A	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0±0	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
C	0.1±0.1	100±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0±0	99.6±0.1	0.1±0.1	100±0.1	100±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1
G	100±0.1	0.1±0.1	100±0.1	0±0	100±0.1	100±0.1	0±0	0±0	0±0	0±0	0.1±0.1	100±0.1	0.1±0.1	99.9±0.1	0±0	0±0	0.1±0.1	0.4±0.1	0±0	0±0	100±0.1	0±0	99.7±0.2	0.1±0.1	100±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.7±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.4±0.2	0±0	0.1±0.1

FIGURE 34 (CONTINUED)

TP53	C-50	C-49	T-48	G-47	T-46	G-45	C-44	A-43	G-42	C-41	T-40	G-39	T-38	G-37	G-36	G-35	T-34	T-33	G-32	A-31	T-30	T-29	C-28	C-27	A-26
Untreated																									
A	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3 + on-target sgRNA																									
A	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	99.8±0.3	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.3±0.3	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3 + off-target sgRNA																									
A	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1

TP53	C-25	A-24	C-23	C-22	C-21	C-20	C-19	G-18	C-17	C-16	C-15	G-14	G-13	C-12	A-11	C-10	C-9	C-8	G-7	C-6	G-5	T-4	C-3	C-2	G-1
Untreated																									
A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3 + on-target sgRNA																									
A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	99.9±0.2	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3 + off-target sgRNA																									
A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.5±0.7	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	99.6±0.6	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1

FIGURE 35

TP53	A1	G2	T3	C4	A5	C6	A7	G8	C9	A10	C11	A12	T13	G14	A15	C16	G17	G18	A19	G20	G21	T22	T23	G24	T25
Untreated																									
A	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
G	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1
BE3 + on-target sgRNA																									
A	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.3±0.4	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.5±0.7	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
G	0.1±0.1	99.9±0.3	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	99.8±0.4	0.1±0.1	100±0.1	99.6±0.7	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1
BE3 + off-target sgRNA																									
A	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
G	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1

TP53	G26	A27	G28	G29	C30	G31	C32	T33	G34	C35	C36	C37	C38	C39	A40	C41	C42	A43	T44	G45	A46	G47	C48	G49	C50
Untreated																									
A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1
G	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3 + on-target sgRNA																									
A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.3	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	99.9±0.3	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1
G	100±0.1	0.1±0.1	99.7±0.6	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	0.4±0.5	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3 + off-target sgRNA																									
A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1
G	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1

FIGURE 35 (CONTINUED)

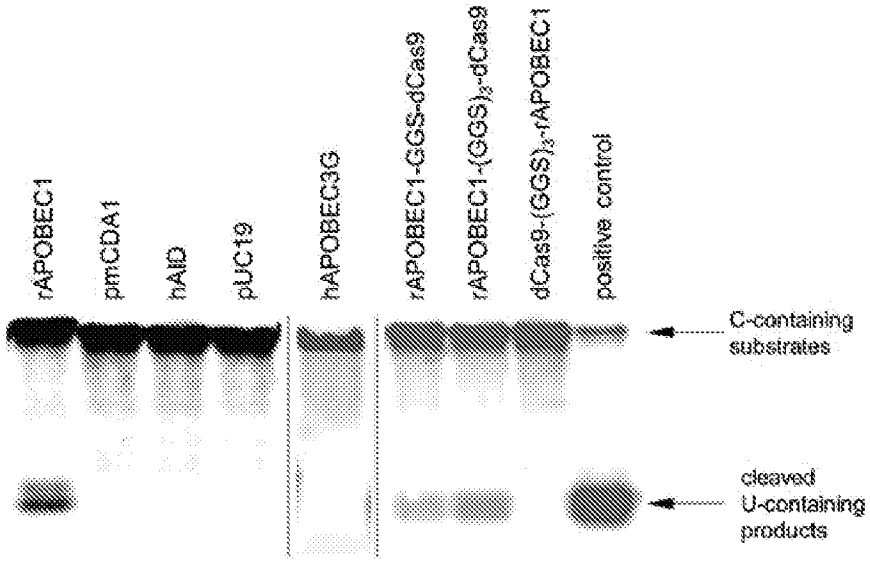


FIGURE 36A

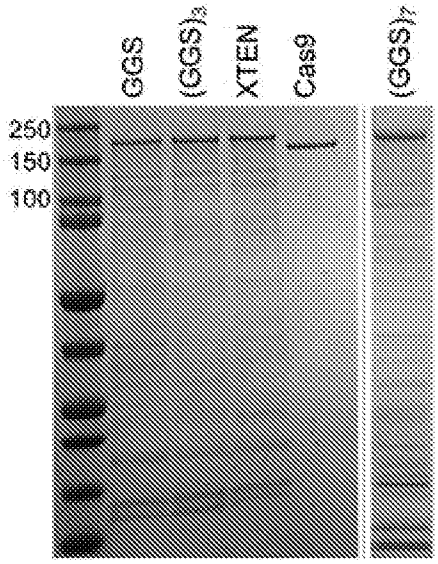


FIGURE 36B

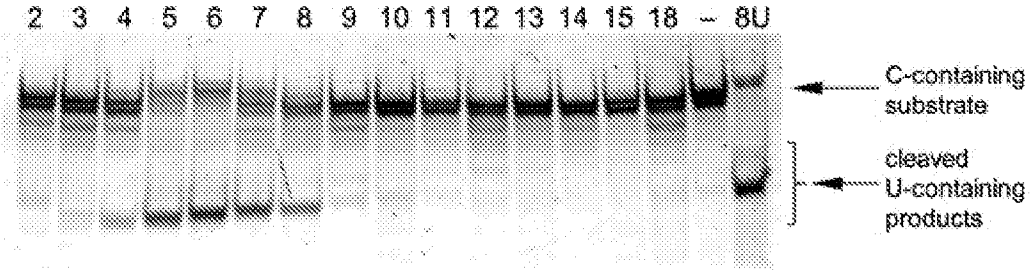


FIGURE 36C

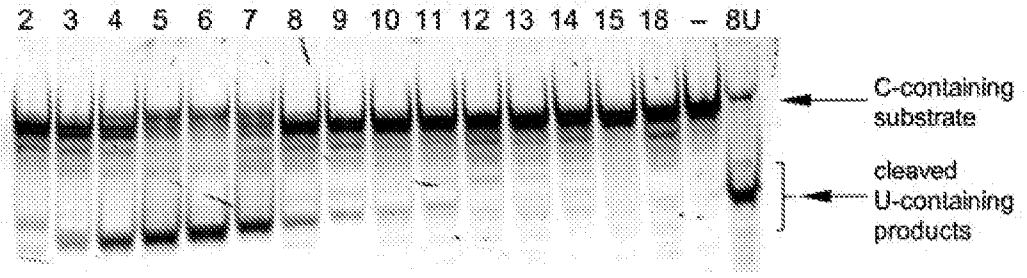


FIGURE 36D

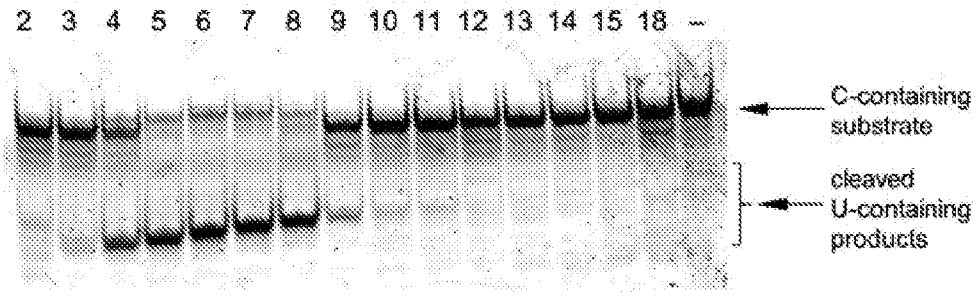


FIGURE 36E

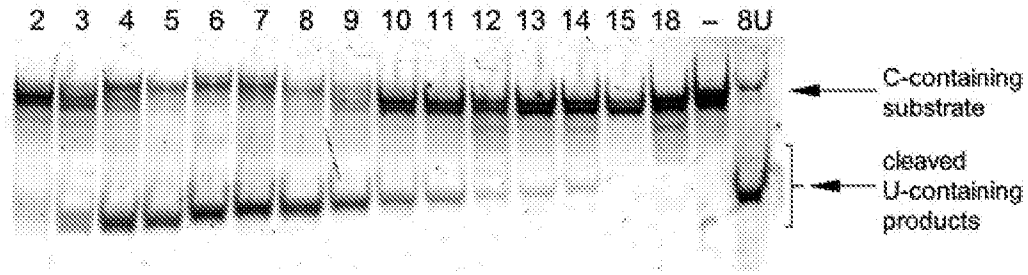


FIGURE 36F

EMX1: GAGTC_5C_6GAGCAGAAGAAGAAGGG

FANCF: GGAAT$C_6C_7C_8$TTC_{11}TGCAGCACCTGG

HEK293 site 2: GAAC_4AC_6AAAGCATAGACTGCGGG

HEK293 site 3: GGCC_4C_5AGACTGAGCACGTGATGG

HEK293 site 4: GGCAC_5TGCGGCTGGAGGTCCGGG

RNF2: GTC_3ATC_6TTAGTC_{12}ATTACCTGAGG

FIGURE 37A

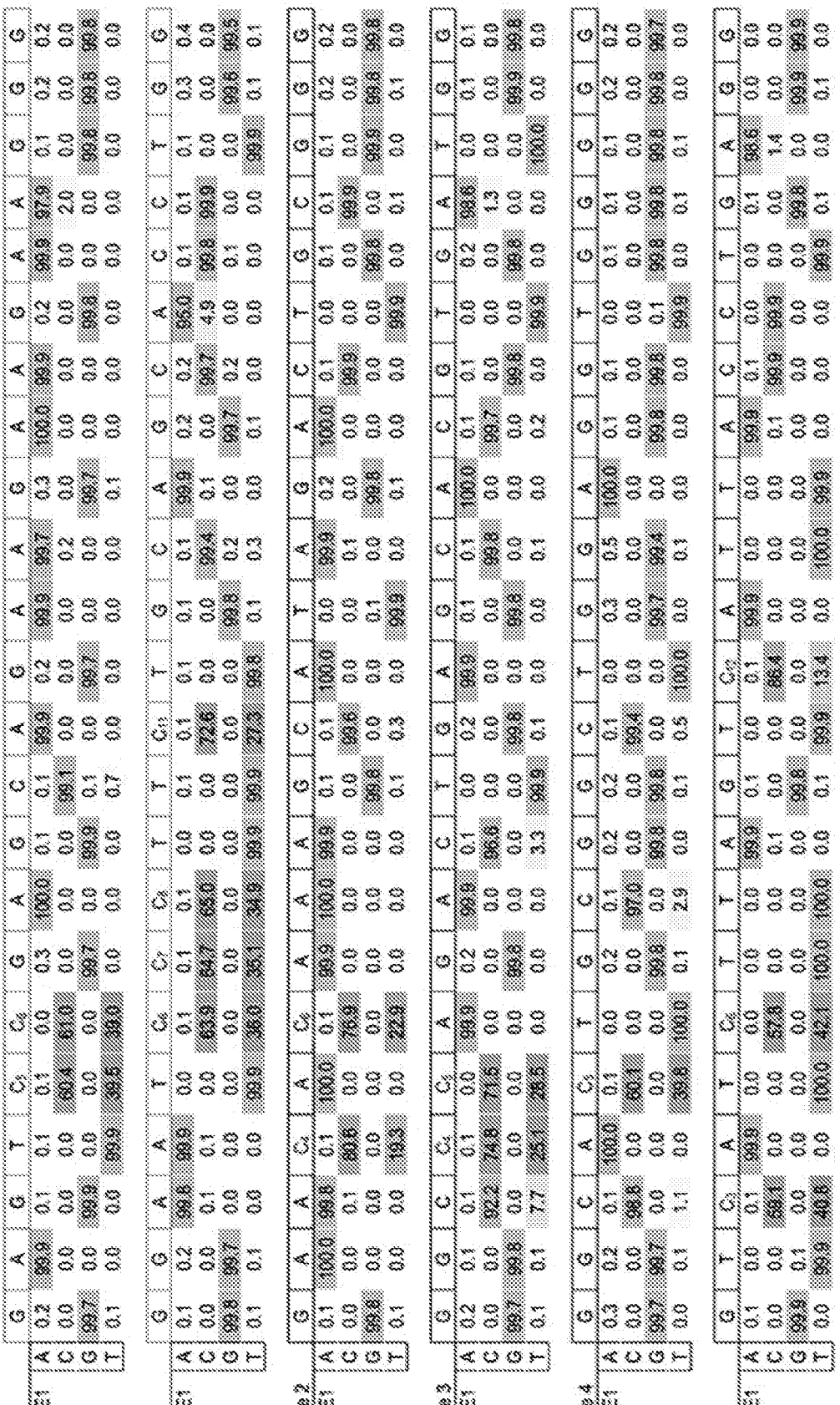

EMX1
1.85 uM BE1
FANCF
1.85 uM BE1
HEK293 site 2
1.85 uM BE1
HEK293 site 3
1.85 uM BE1
HEK293 site 4
1.85 uM BE1
RNF2
1.85 uM BE1

FIGURE 37B

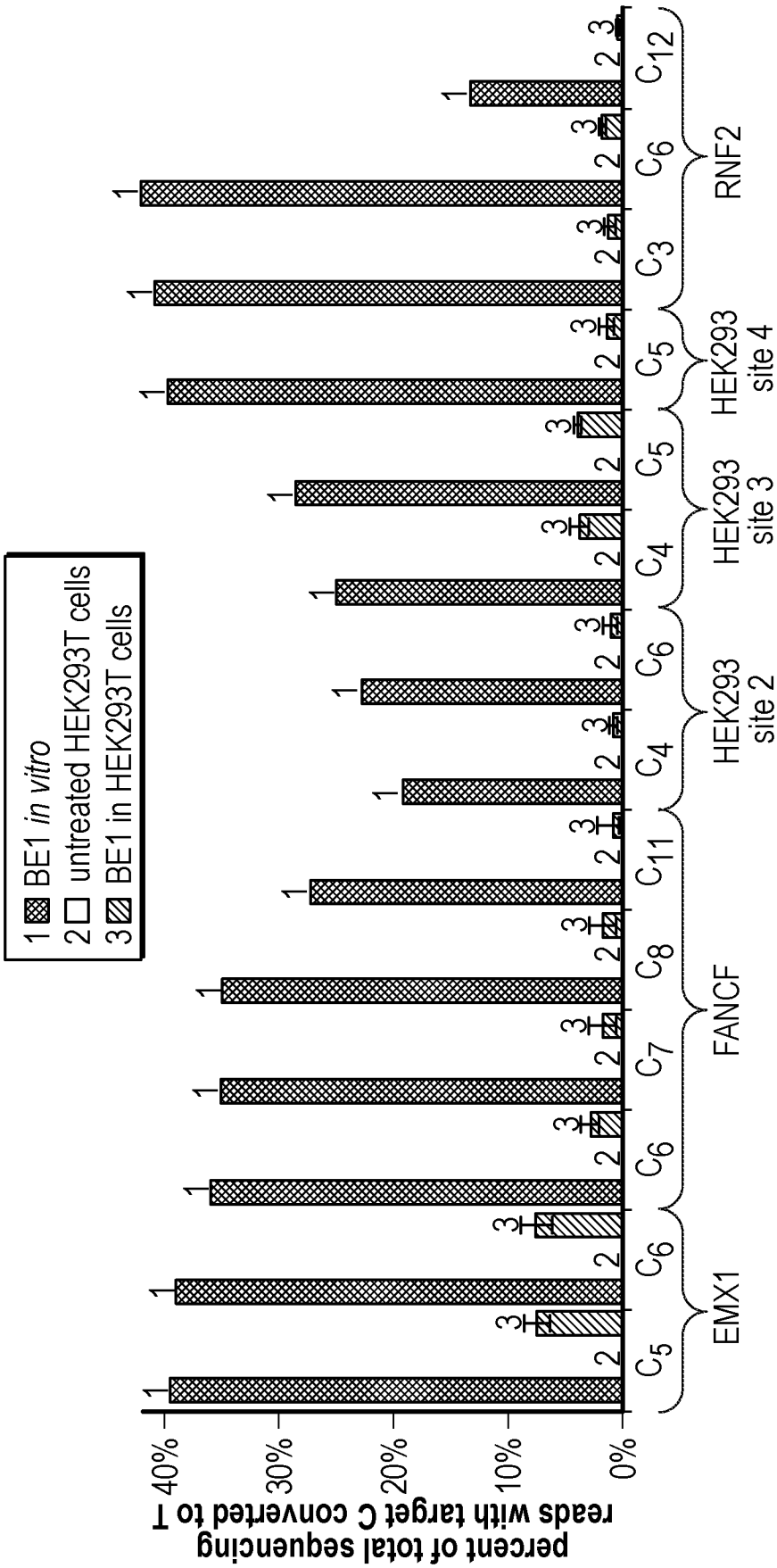

1 BE1 in vitro
2 untreated HEK293T cells
3 BE1 in HEK293T cells
percent of total sequencing reads with target C converted to T
40%
30%
20%
10%
0%
EMX1
FANCF
HEK293 site 2
HEK293 site 3
HEK293 site 4
RNF2
C5
C6
C7
C8
C11
C12

FIGURE 37C

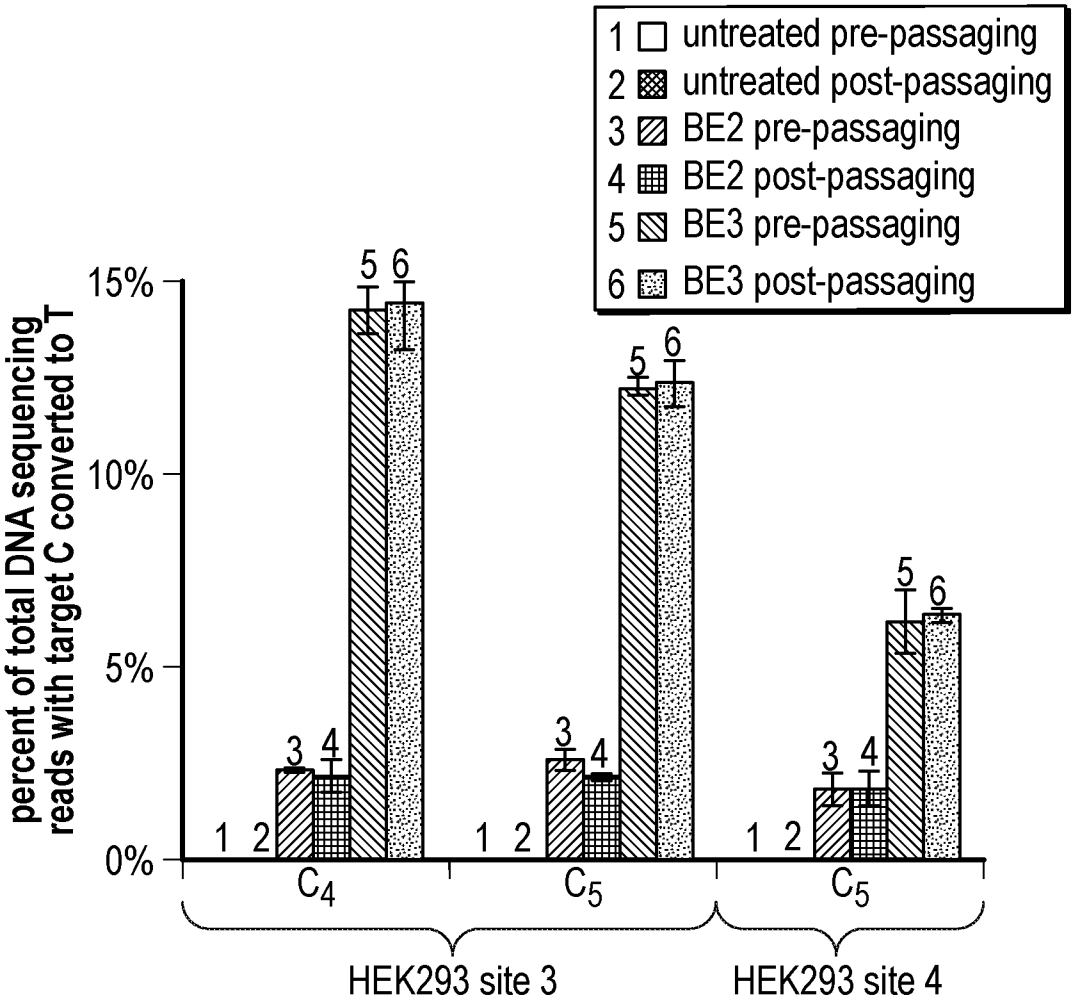

1 untreated pre-passaging
2 untreated post-passaging
3 BE2 pre-passaging
4 BE2 post-passaging
5 BE3 pre-passaging
6 BE3 post-passaging
percent of total DNA sequencing reads with target C converted to T
15%
10%
5%
0%
C4
C5
C5
HEK293 site 3
HEK293 site 4

FIGURE 38

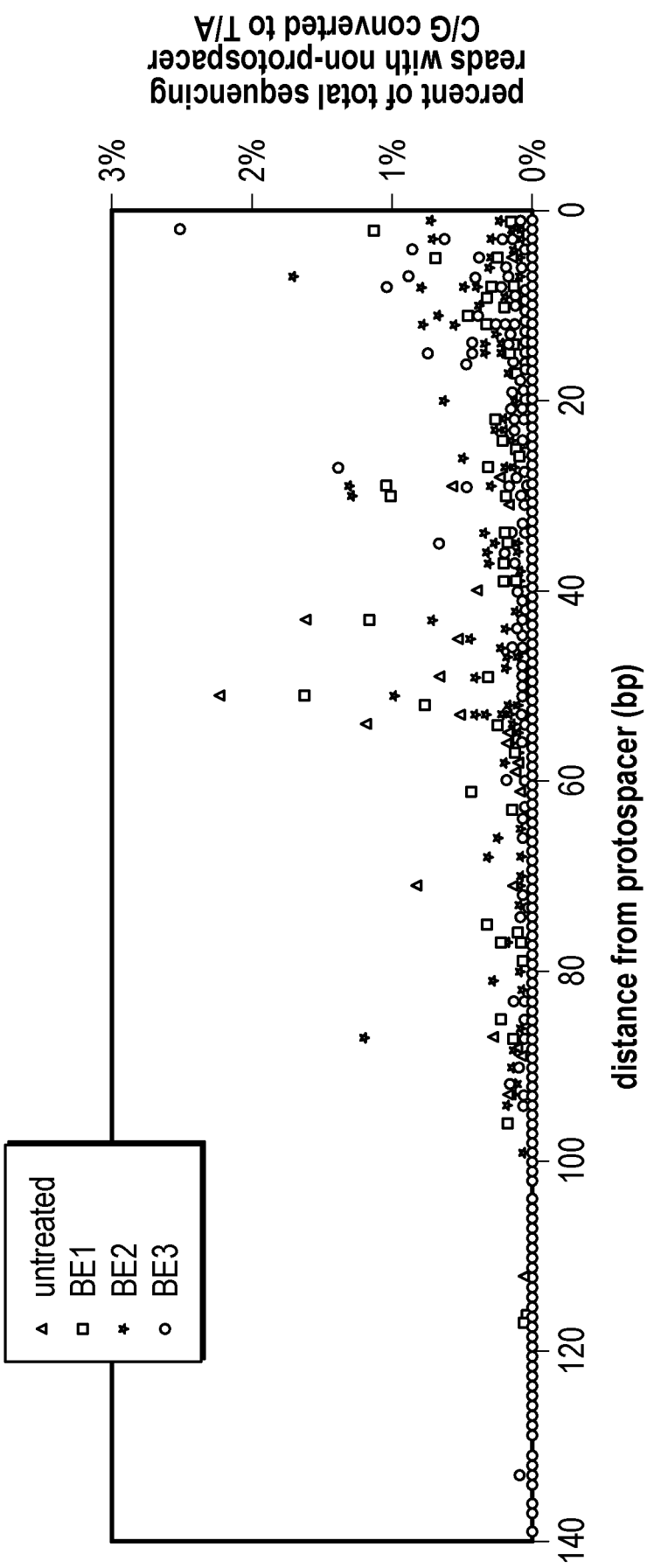

percent of total sequencing reads with non-protospacer C/G converted to T/A
3%
2%
1%
0%
0
20
40
60
80
100
120
140
distance from protospacer (bp)
untreated
BE1
BE2
BE3

FIGURE 39A

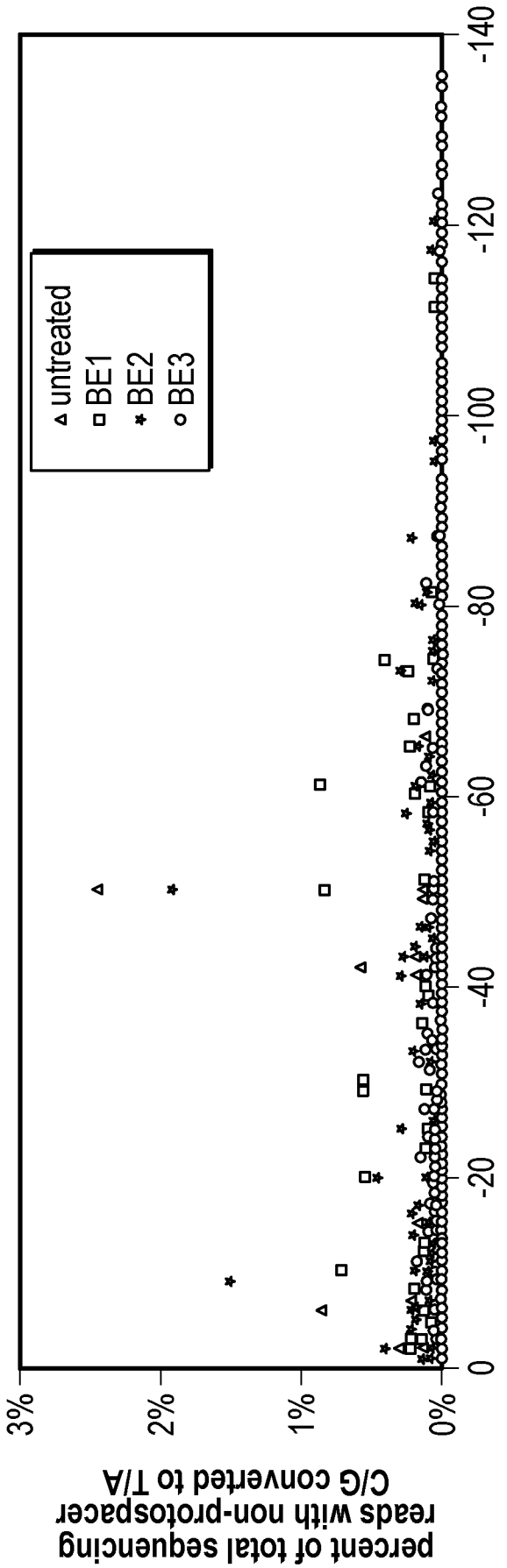

percent of total sequencing reads with non-protospacer C/G converted to T/A
3%
2%
1%
0%
0
-20
-40
-60
-80
-100
-120
-140
distance from protospacer (bp)
untreated
BE1
BE2
BE3

FIGURE 39B

non-protospacer C/Gs	average C/G (%)	average T/A (%)	lowest T/A (%)	highest T/A (%)
untreated	99.95 ± 0.14	0.02 ± 0.02	0.00	2.44
BE1	99.95 ± 0.24	0.03 ± 0.03	0.00	1.64
BE2	99.95 ± 0.13	0.03 ± 0.03	0.00	1.92
BE3	99.97 ± 0.09	0.02 ± 0.02	0.00	2.52

FIGURE 39C

untreated
APOE4 C158R
BE3 + on-target sgRNA
APOE4 C158R
BE3 + off-target sgRNA
APOE4 C158R
Cas9 + HDR
APOE4 C158R
Lys
Arg
Leu
Ala
Val
Tyr
Gln
indel %
Arg → Cys
Leu → Leu

FIGURE 40A

untreated
TP53 Y163C
BE3 + on-target sgRNA
TP53 Y163C
BE3 + off-target sgRNA
TP53 Y163C
Cas9 + HDR
TP53 Y163C
Arg
Ala
Met
Ile
Cys
Cys → Tyr
Lys
indel %

FIGURE 40B

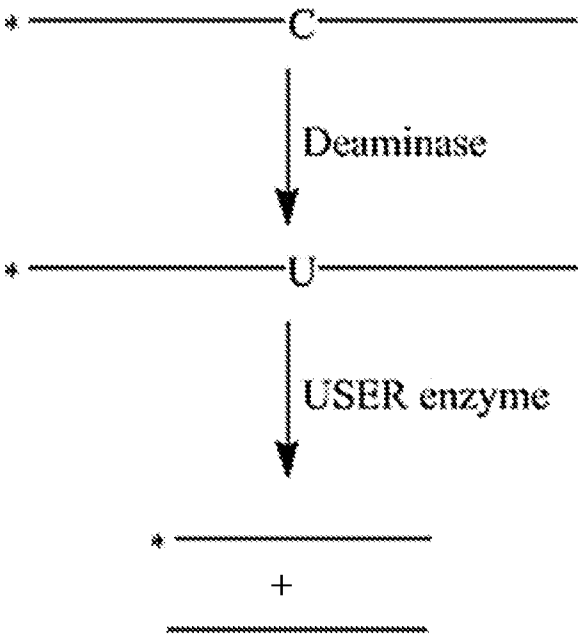


FIGURE 41

pmCDA1
nCas9
UGI
NLS
T G C C A T C A C G T G C T C A A T C G G G C C C A
P A M 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 -1 -2 -3
CDA
A 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 99.9% 0.0% 0.0% 0.0% 0.5% 0.0% 0.0% 0.1% 100.0% 5.8% 0.0% 0.0% 0.0% 0.0% 33.9% 33.9% 33.1% 0.0% 0.0% 0.0% 0.0% 100.0%
C 0.0% 0.0% 99.9% 99.9% 0.0% 0.0% 99.9% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 99.9% 0.0% 1.0% 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 100.0% 99.9% 99.9% 0.0%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 99.5% 0.0% 0.0% 0.0% 0.0% 91.8% 0.0% 0.0% 0.0% 65.7% 65.0% 65.2% 0.0% 0.0% 0.0% 0.0% 0.0%
T 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 1.5% 100.0% 0.0% 100.0% 0.3% 1.1% 1.7% 0.0% 0.0% 0.0% 0.0% 0.0%
rAPOBEC1
A 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.1% 100.0% 0.8% 0.0% 0.0% 0.0% 38.5% 41.9% 1.4% 0.0% 0.0% 0.1% 0.0% 100.0%
C 0.0% 0.0% 99.9% 99.9% 0.0% 0.0% 99.9% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 99.9% 0.0% 0.1% 0.0% 99.9% 0.0% 8.1% 0.7% 0.0% 100.0% 99.9% 99.9% 100.0% 0.0%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 99.9% 0.0% 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 99.0% 0.0% 0.0% 0.0% 51.3% 56.7% 98.6% 0.0% 0.0% 0.0% 0.0% 0.0%
T 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.1% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 100.0% 2.2% 0.7% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%
untreated
A 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.1% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.1% 100.0% 0.0% 0.0% 0.1% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0%
C 0.0% 0.0% 99.9% 99.9% 0.0% 0.0% 99.9% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 99.9% 0.0% 0.0% 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 100.0% 100.0% 99.9% 100.0% 0.0%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 99.9% 99.9% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0%
T 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%

FIGURE 42

pmCDA1
nCas9
UGI
NLS
Hek-2 site, shown: coding strand (G→A), target: template stand (C→T)
CDA
rAPOBEC1
untreated

FIGURE 43

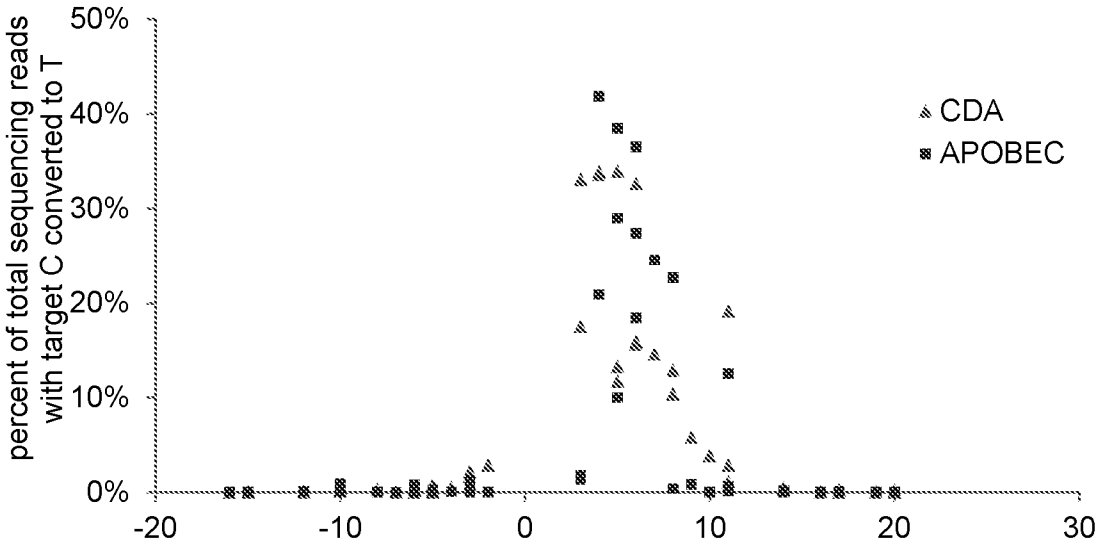


FIGURE 44

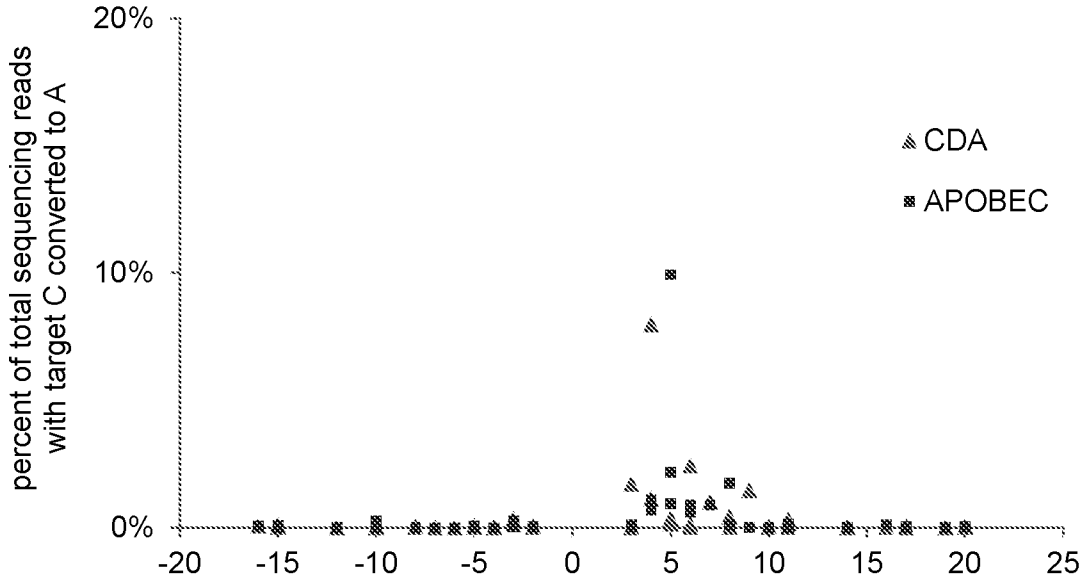


FIGURE 45

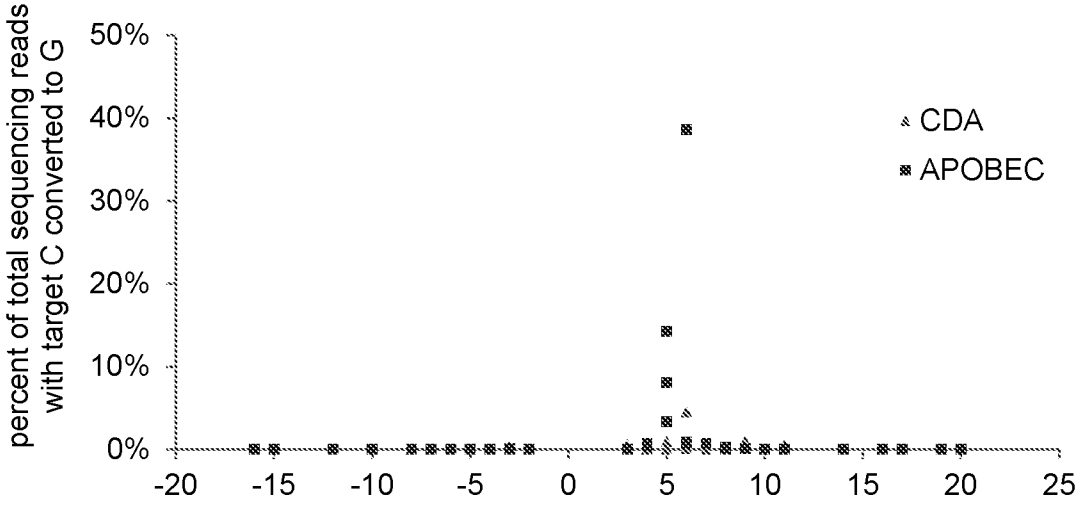
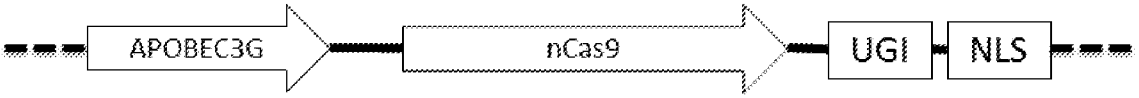
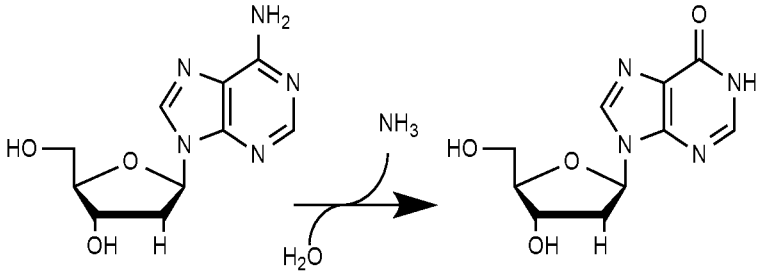

percent of total sequencing reads with target C converted to G
50%
40%
30%
20%
10%
0%
-20
-15
-10
-5
0
5
10
15
20
25
CDA
APOBEC

FIGURE 46



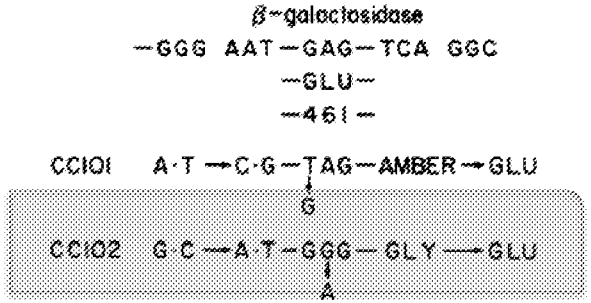
	P	A	M	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
	C	C	C	G	C	A	G	T	C	T	A	T	G	C	T	T	T	G	T	G	T	T	C
rAPOBEC1																							
A	0.02%	0.02%	0.01%	0.02%	0.02%	0.02%	0.05%	0.01%	99.96%	0.09%	0.00%	0.01%	0.00%	99.96%	0.01%	0.02%	0.01%	0.00%	0.00%	0.00%	9.30%	0.01%	10.87%	0.00%	0.01%	0.01%
C	99.96%	0.00%	99.98%	99.97%	99.97%	99.97%	0.00%	99.98%	0.02%	0.01%	0.01%	99.98%	0.01%	0.01%	0.01%	0.01%	99.98%	0.01%	0.01%	0.01%	18.14%	0.01%	0.54%	0.01%	0.01%	99.95
G	0.00%	99.96%	0.00%	0.00%	0.00%	0.00%	99.93%	0.00%	0.01%	99.86%	0.00%	0.00%	0.00%	0.01%	0.00%	99.96%	0.00%	0.00%	0.00%	0.00%	72.40%	0.00%	88.10%	0.00%	0.00%	0.01%
T	0.02%	0.02%	0.01%	0.01%	0.01%	0.01%	0.02%	0.01%	0.01%	0.03%	99.99%	0.01%	99.99%	0.02%	99.98%	0.01%	0.01%	99.99%	99.98%	99.99%	0.17%	99.98%	0.49%	99.98%	99.98%	0.03%
APOBEC3G	0.01%	0.01%	0.01%	0.01%	0.02%	0.01%	0.05%	0.01%	99.97%	0.01%	0.00%	0.00%	0.00%	99.97%	0.01%	0.01%	0.00%	0.00%	0.01%	0.00%	2.18%	0.01%	0.34%	0.00%	0.00%	0.01%
A	99.97%	0.01%	99.98%	99.98%	99.96%	99.98%	0.00%	99.99%	0.02%	0.00%	0.01%	99.99%	0.01%	0.01%	0.01%	0.00%	99.98%	0.01%	0.01%	0.01%	5.34%	0.01%	0.15%	0.01%	0.02%	99.98
C	0.00%	99.97%	0.00%	0.00%	0.00%	0.00%	99.93%	0.00%	0.01%	99.98%	0.00%	0.00%	0.00%	0.00%	0.00%	99.98%	0.00%	0.00%	0.00%	0.00%	92.44%	0.00%	99.46%	0.00%	0.00%	0.00%
G	0.02%	0.01%	0.01%	0.01%	0.02%	0.01%	0.02%	0.00%	0.00%	0.01%	99.99%	0.01%	99.99%	0.02%	99.98%	0.01%	0.01%	99.99%	99.98%	99.99%	0.04%	99.98%	0.05%	99.98%	99.98%	0.01%
T																										
APOBEC3G_RR	0.01%	0.02%	0.01%	0.01%	0.02%	0.01%	0.05%	0.01%	99.97%	0.01%	0.00%	0.00%	0.00%	99.97%	0.01%	0.04%	0.01%	0.00%	0.01%	0.00%	2.18%	0.01%	6.12%	0.00%	0.00%	0.01%
A	99.97%	0.00%	99.98%	99.99%	99.97%	99.97%	0.00%	99.98%	0.02%	0.00%	0.01%	99.99%	0.01%	0.01%	0.01%	0.01%	99.98%	0.01%	0.01%	0.01%	4.98%	0.01%	2.49%	0.01%	0.02%	99.99
C	0.00%	99.96%	0.00%	0.00%	0.00%	0.00%	99.92%	0.00%	0.01%	99.98%	0.00%	0.00%	0.00%	0.01%	0.00%	99.94%	0.00%	0.00%	0.00%	0.00%	92.80%	0.00%	91.04%	0.00%	0.00%	0.00%
G	0.02%	0.02%	0.01%	0.01%	0.01%	0.02%	0.02%	0.01%	0.00%	0.01%	99.99%	0.01%	99.99%	0.01%	99.98%	0.01%	0.01%	99.99%	99.99%	99.98%	0.05%	99.98%	0.35%	99.98%	99.98%	0.01%
T																										
untreated																										
A	0.02%	0.02%	0.00%	0.01%	0.02%	0.01%	0.05%	0.01%	99.98%	0.00%	0.00%	0.01%	0.00%	99.97%	0.01%	0.00%	0.01%	0.00%	0.00%	0.00%	0.01%	0.01%	0.01%	0.00%	0.01%	0.01%
C	99.96%	0.00%	99.99%	99.98%	99.97%	99.97%	0.00%	99.98%	0.01%	0.00%	0.01%	99.99%	0.01%	0.01%	0.01%	0.00%	99.98%	0.01%	0.01%	0.01%	0.01%	0.01%	0.00%	0.01%	0.01%	99.98
G	0.00%	99.96%	0.00%	0.00%	0.00%	0.00%	99.91%	0.00%	0.01%	99.98%	0.00%	0.00%	0.00%	0.00%	0.00%	99.98%	0.00%	0.00%	0.00%	0.00%	99.96%	0.00%	99.96%	0.00%	0.00%	0.00%
T	0.02%	0.02%	0.01%	0.01%	0.01%	0.02%	0.03%	0.01%	0.00%	0.01%	99.99%	0.01%	99.99%	0.02%	99.98%	0.01%	0.01%	99.99%	99.99%	99.99%	0.02%	99.98%	0.02%	99.98%	99.98%	0.01%

FIGURE 47



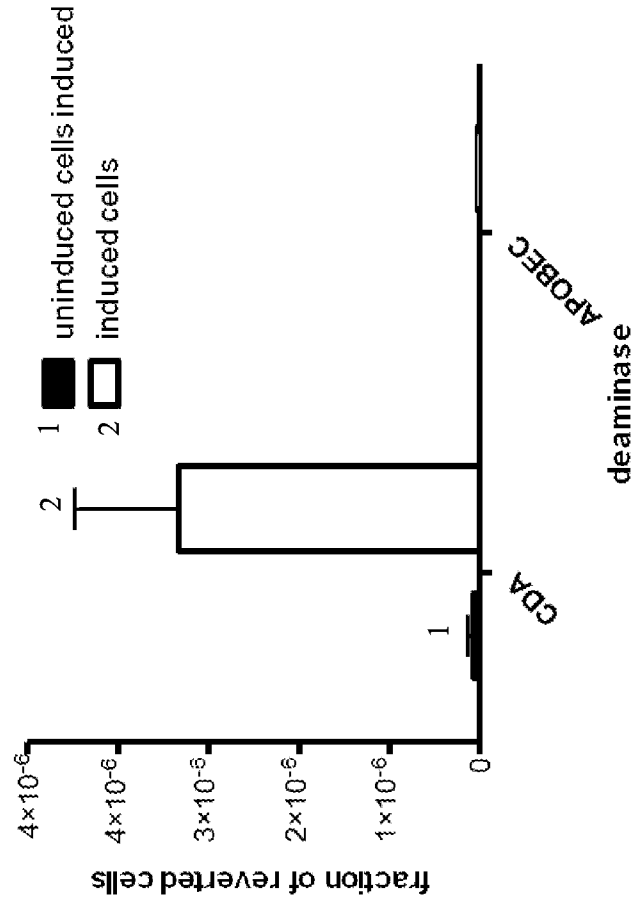
adenosine inosine

Utilizing a bacterial strain with lacZ encoded on the F plasmid
Critical glutamic acid residue is mutated. Reversion restores lacZ activity.



Strain used to develop selection assay
Constructs used: APOBEC1 and CDA

FIGURE 48


fraction of reverted cells
4×10-6
4×10-6
3×10-6
2×10-6
1×10-6
0
1
2
CDA
APOBEC
deaminase
1 uninduced cells induced
2 induced cells

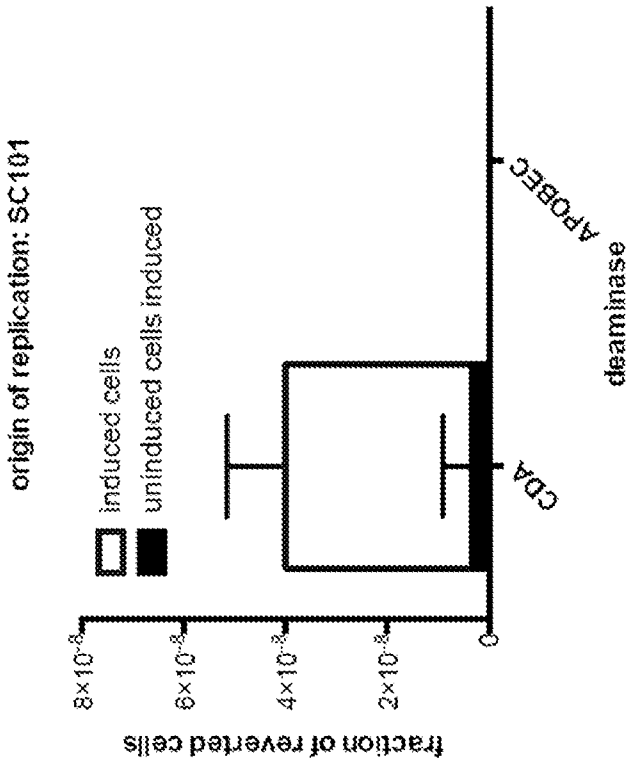

origin of replication: SC101
induced cells
uninduced cells induced
fraction of reverted cells
8×10-8
6×10-8
4×10-8
2×10-8
0
CDA
APOBEC
deaminase

FIGURE 49

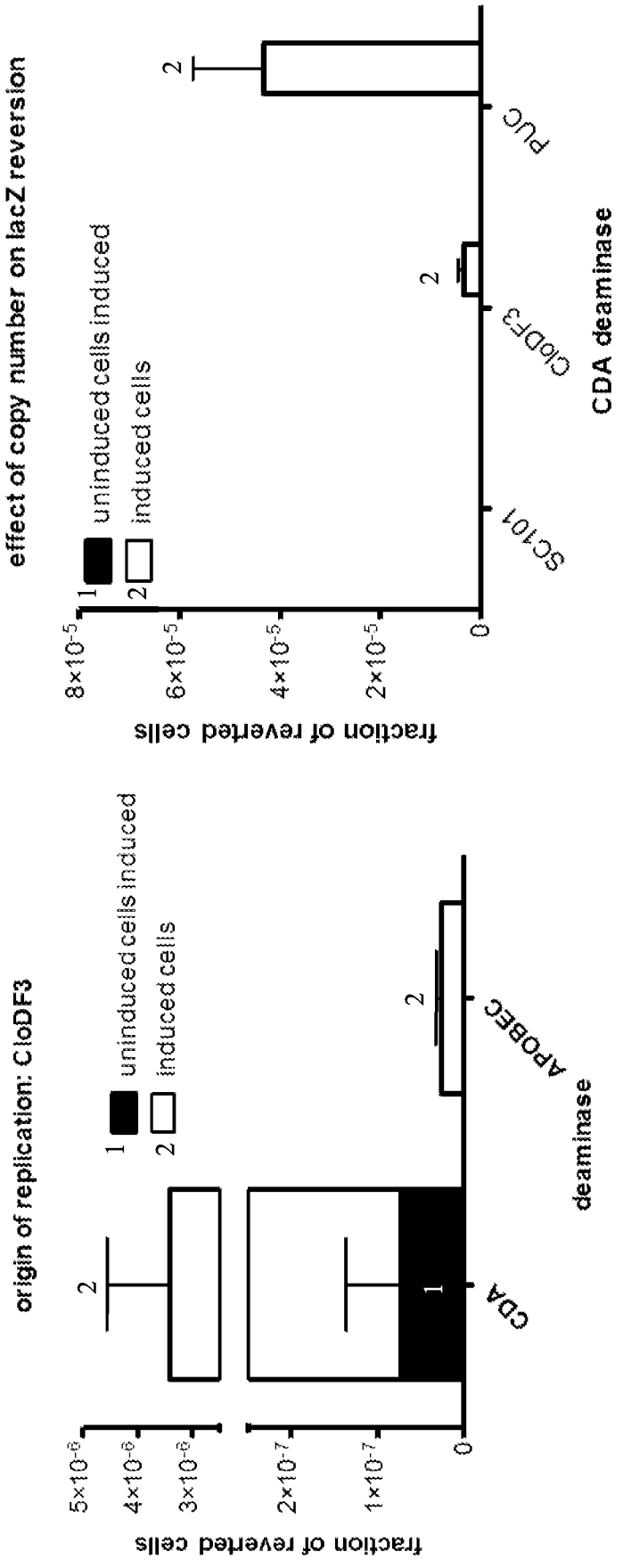


FIGURE 49 (CONTINUED)

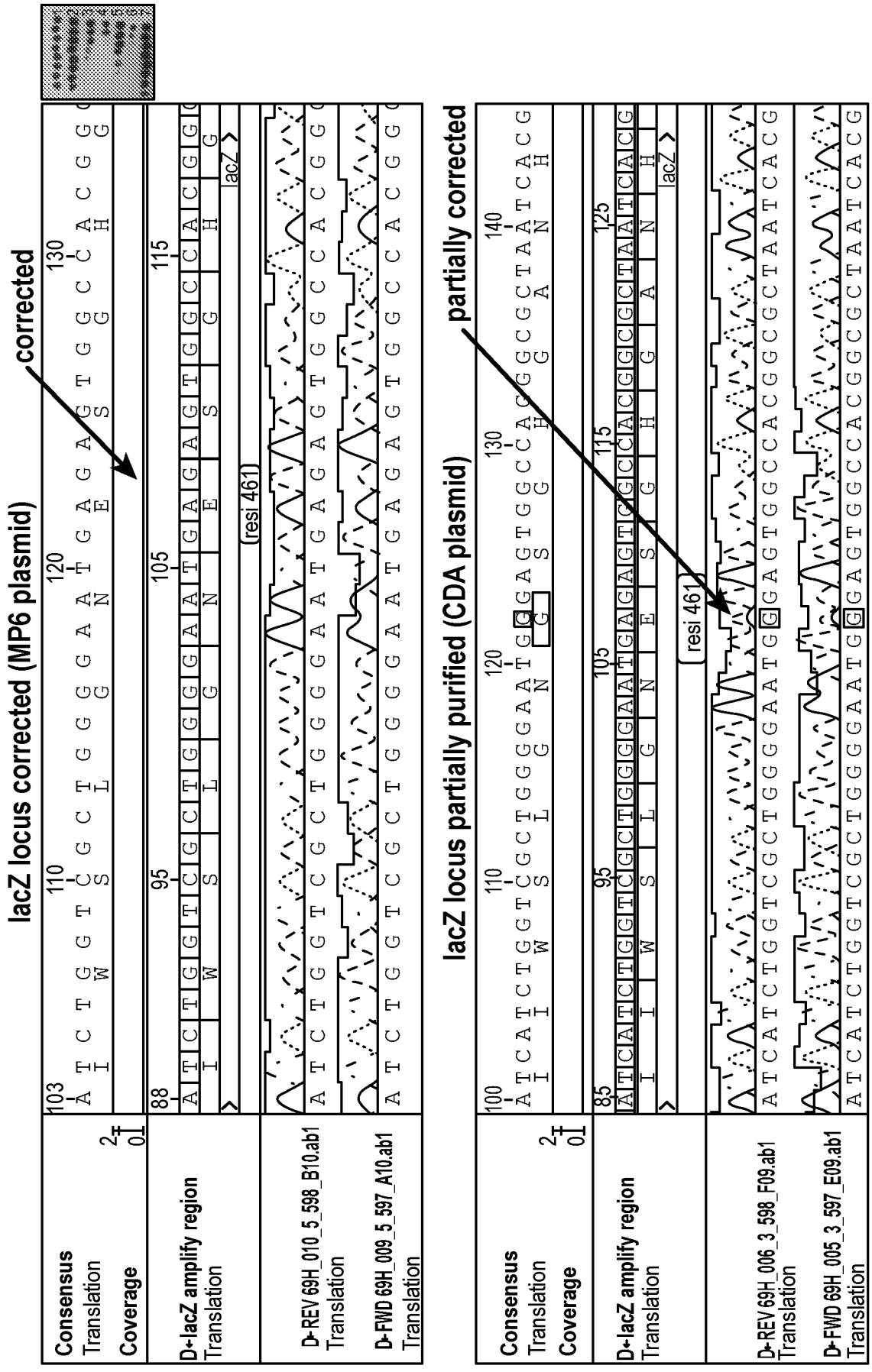

lacZ locus corrected (MP6 plasmid)
corrected
Consensus
Translation
Coverage
D-lacZ amplify region
Translation
D-REV 69H_010_5_598_B10.ab1
Translation
D-FWD 69H_009_5_597_A10.ab1
Translation
(resi 461)
lacZ
lacZ locus partially purified (CDA plasmid)
partially corrected
D-REV 69H_006_3_598_F09.ab1
Translation
D-FWD 69H_005_3_597_E09.ab1
Translation
resi 461

FIGURE 50

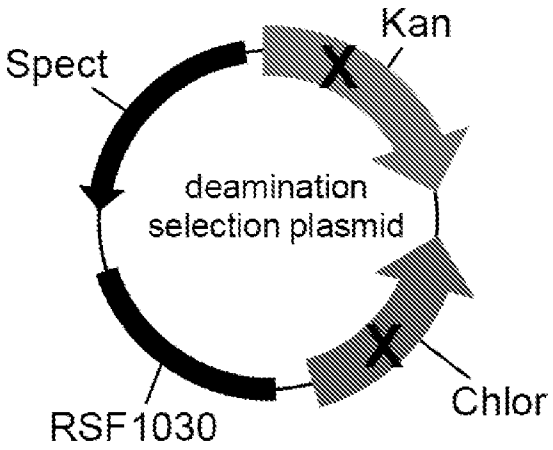
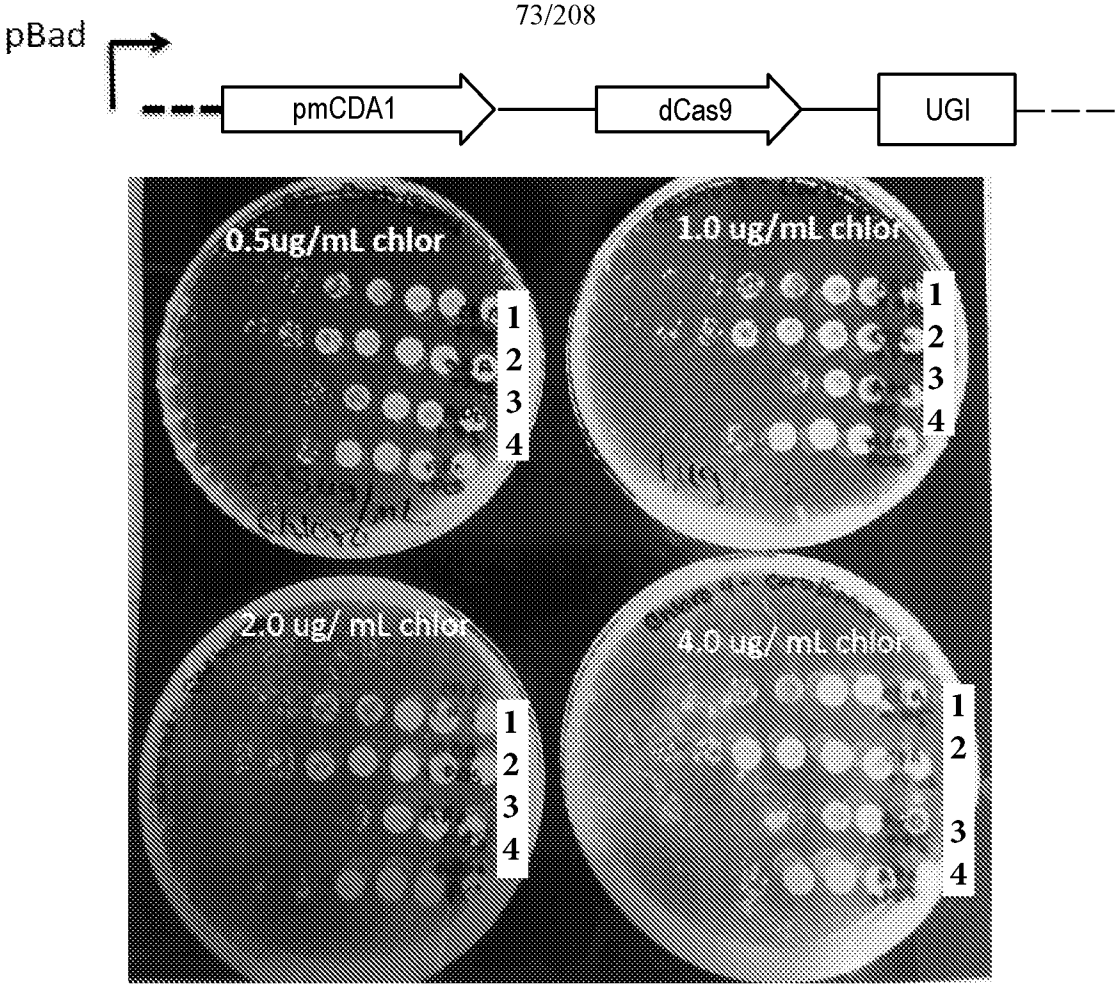


FIGURE 51



Row 1: CDA-dCas9 + selection plasmid ($chlor^S$)

Row 2: CDA-dCas9 + pos. control selection ($chlor^R$)

Row 3: rAPOBEC-dCas9 + selection plasmid ($chlor^S$)

Row 4: rAPOBEC-dCas9 + pos. control selection ($chlor^R$)

FIGURE 52

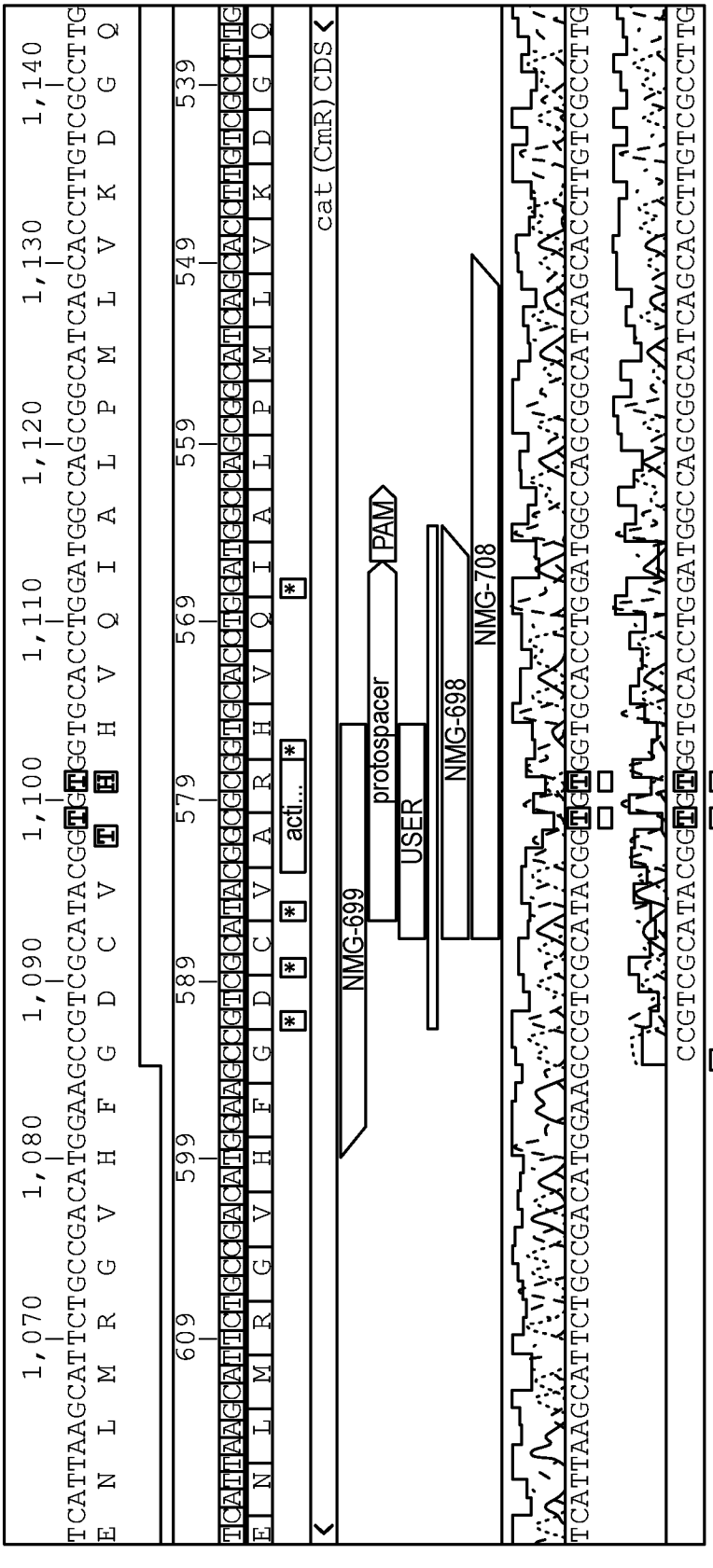

rAPOBEC-XTEN-dCas9-UGI on 8 ug/mL chlor
cat(CmR)CDS
NMG-699
protospacer
PAM
USER
NMG-698
NMG-708

FIGURE 53A

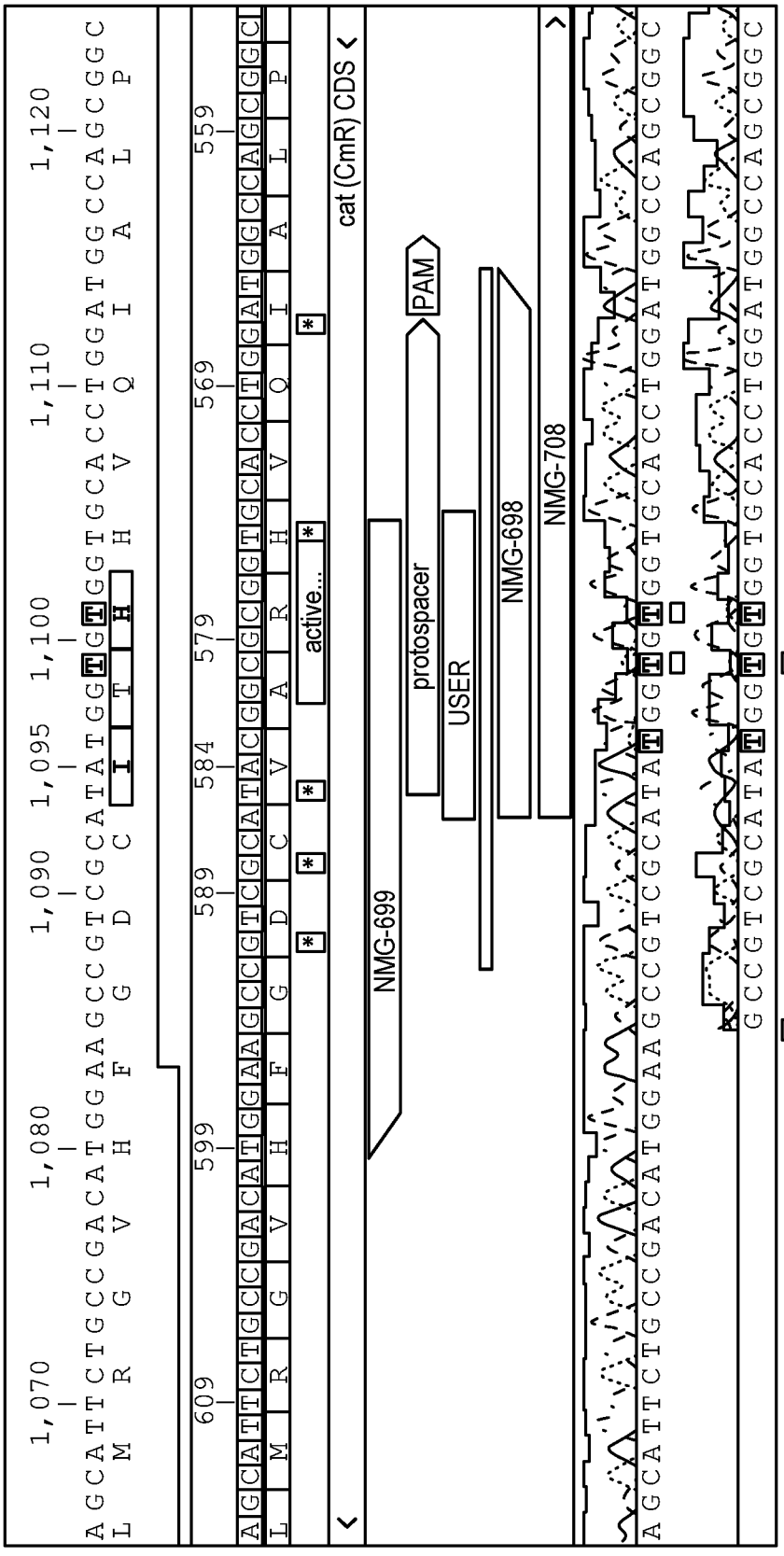

CDA-XTEN-dCas9-UGI: survival on 8 ug/mL chloramphenicol
cat (CmR) CDS
NMG-699
protospacer
PAM
USER
NMG-698
NMG-708

FIGURE 53B

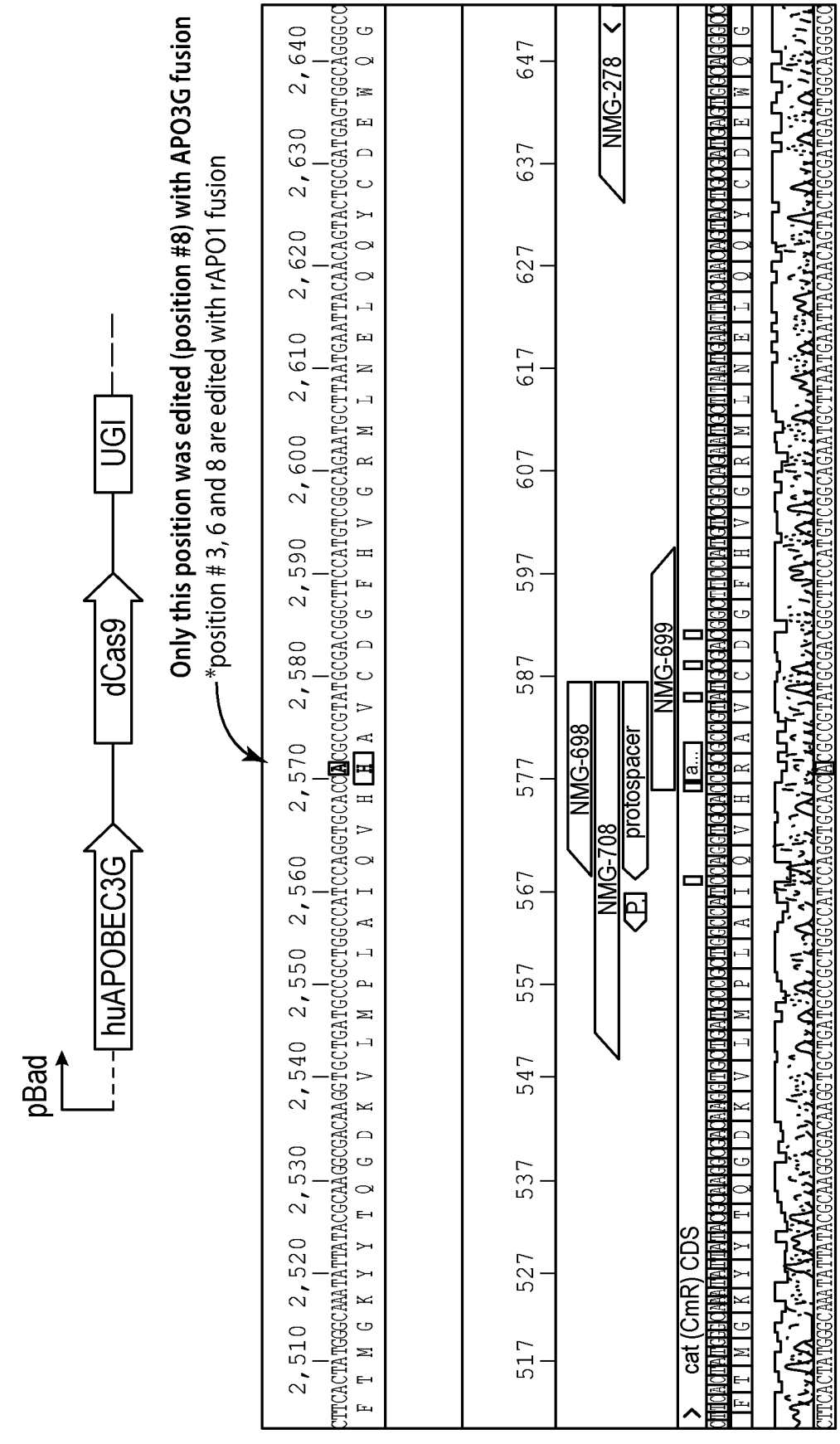

pBad
huAPOBEC3G
dCas9
UGI
Only this position was edited (position #8) with APO3G fusion
*position # 3, 6 and 8 are edited with rAPO1 fusion
NMG-278
NMG-698
NMG-708
protospacer
NMG-699
cat (CmR) CDS

FIGURE 54

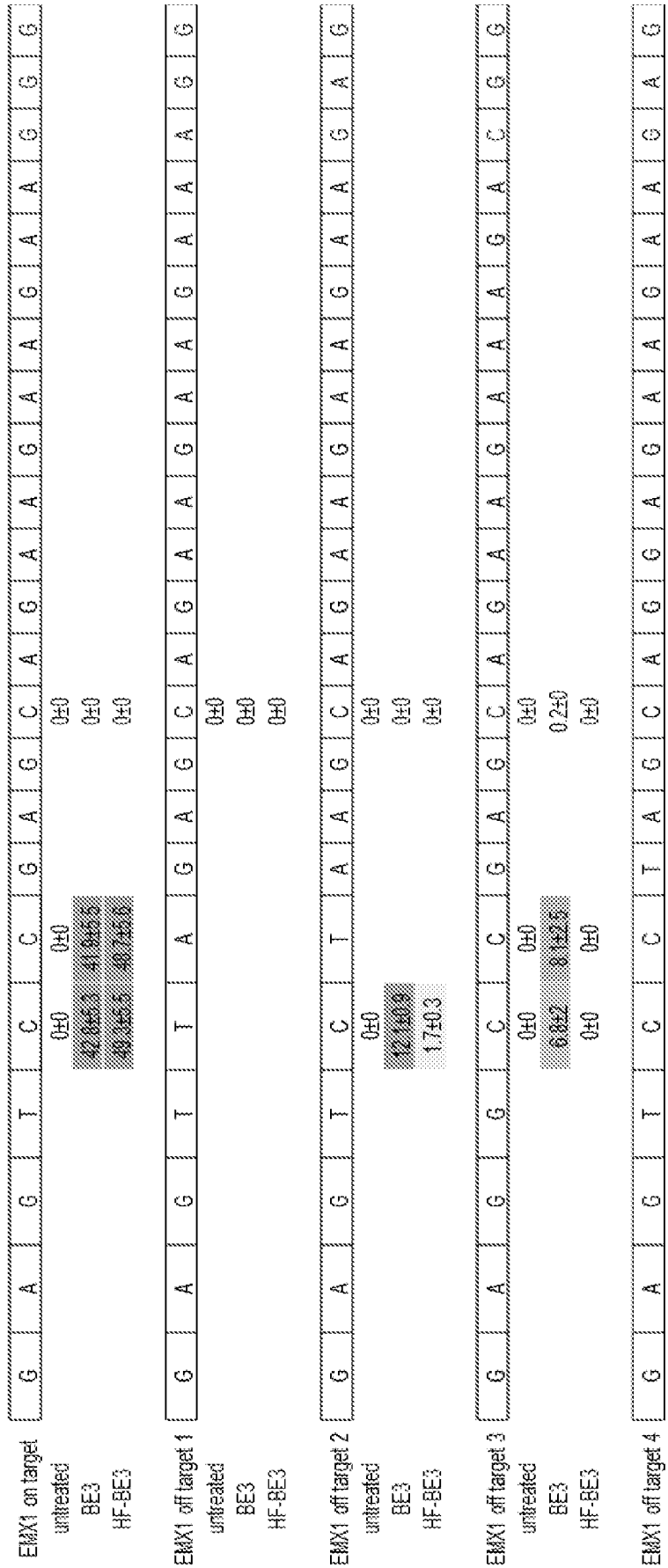

EMX1 on target
untreated
BE3
HF-BE3
EMX1 off target 1
untreated
BE3
HF-BE3
EMX1 off target 2
untreated
BE3
HF-BE3
EMX1 off target 3
untreated
BE3
HF-BE3
EMX1 off target 4
untreated
BE3
HF-BE3

FIGURE 55

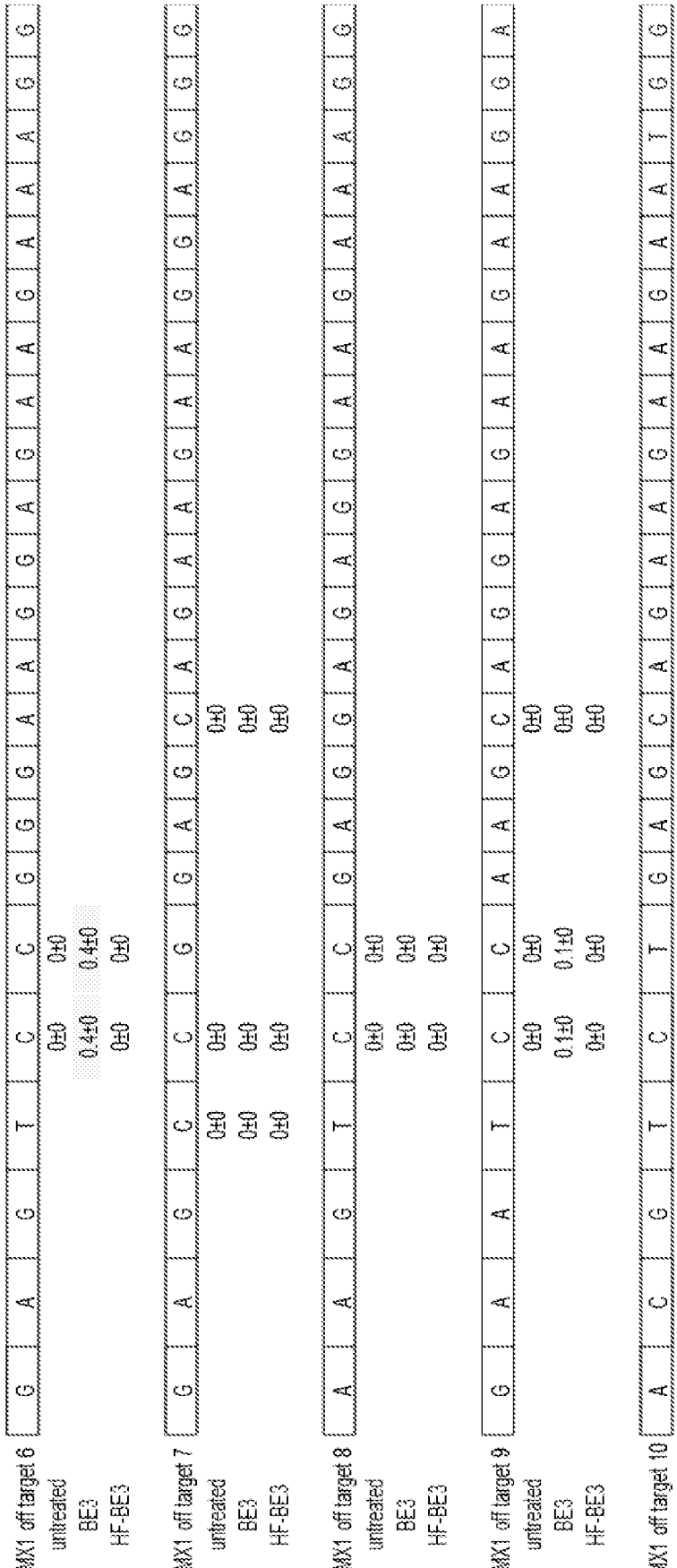


FIGURE 55 (CONTINUED)

EMX1		C_3	C_6	C_{12}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	
BE3	A	1.3 ± 0.1	0.7 ± 0.2	0 ± 0	2.6 ± 0.3
	C	51.7 ± 1	55.2 ± 0.1	99.9 ± 0	
	G	4 ± 0.7	1.9 ± 0.4	0 ± 0	
	T	42.8 ± 0.2	41.9 ± 0.4	0 ± 0	
HF BE3	A	1.5 ± 0.7	0.5 ± 0.1	0 ± 0	1 ± 0.1
	C	44.8 ± 8.5	49.9 ± 6.6	99.9 ± 0	
	G	4.2 ± 1.2	0.6 ± 0.2	0 ± 0	
	T	49.3 ± 6.5	48.7 ± 6.3	0 ± 0	

FANCF		C_6	C_7	C_8	C_{11}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0.8 ± 0.1	1.2 ± 0.3	1.2 ± 0.1	0.2 ± 0.1	5.8 ± 0.9
	C	64.7 ± 5.9	67.9 ± 5.1	68.9 ± 5.2	85.4 ± 2.7	
	G	0.6 ± 0	0.7 ± 0	0.7 ± 0.1	0.1 ± 0	
	T	33.8 ± 5.7	29.9 ± 4.8	29 ± 5	14.1 ± 2.6	
HF BE3	A	0.9 ± 0.2	2 ± 0.4	0.9 ± 0.2	0.4 ± 0	5.9 ± 0.7
	C	52.5 ± 8.4	57.7 ± 7.6	61.9 ± 6.8	84.7 ± 2.4	
	G	1.4 ± 0.2	0.6 ± 0	0.5 ± 0.1	0.3 ± 0	
	T	45 ± 8	39.5 ± 7.2	36.5 ± 6.5	14.4 ± 2.2	

RNF2		C_3	C_6	C_{12}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0.4 ± 0	1.5 ± 0.2	0.1 ± 0	2.3 ± 0.3
	C	70.9 ± 2.7	45 ± 0	95 ± 0.2	
	G	0 ± 0	12.6 ± 2.4	0.1 ± 0	
	T	28.5 ± 2.7	40.7 ± 2.7	4.6 ± 0.2	
HF BE3	A	0 ± 0	0.2 ± 0	0 ± 0	0.5 ± 0.1
	C	76.8 ± 1.8	64.5 ± 3.3	95.4 ± 0.1	
	G	0 ± 0	2.4 ± 0.5	0 ± 0	
	T	23 ± 1.8	32.7 ± 2.6	4.5 ± 0.1	

HEK3		C_6	C_7	C_8	C_{11}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0 ± 0	1.6 ± 0.2	1.5 ± 0.3	0.2 ± 0	2.7 ± 0.4
	C	98.7 ± 0.2	48.6 ± 5.4	40.8 ± 6.7	98.6 ± 0.1	
	G	0 ± 0	1.4 ± 0.2	10.5 ± 1.2	0.3 ± 0	
	T	1.1 ± 0.2	48.2 ± 4.9	47 ± 5.2	0.7 ± 0.1	
HF BE3	A	0 ± 0	0.8 ± 0.2	1.1 ± 0.3	0.4 ± 0	4.1 ± 0.7
	C	97.8 ± 0.4	65 ± 3.8	40.4 ± 6.3	95.9 ± 0.5	
	G	0 ± 0	0.9 ± 0.2	6.1 ± 1.1	0.7 ± 0.2	
	T	2 ± 0.4	33 ± 3.4	52.2 ± 5.2	2.7 ± 0.2	

HEK4		C_6	C_7	C_8	C_{11}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0 ± 0	5.4 ± 1.3	0 ± 0	0 ± 0	3 ± 0.7
	C	98.7 ± 0	44.5 ± 8.2	96.9 ± 0	99.8 ± 0	
	G	0 ± 0	18.1 ± 2.3	0 ± 0	0 ± 0	
	T	1.1 ± 0.1	31.9 ± 4.8	0.9 ± 0.1	0.1 ± 0	
HF BE3	A	0 ± 0	5.2 ± 0.8	0 ± 0	0 ± 0	1.3 ± 0
	C	99.6 ± 0	36.6 ± 9	98.5 ± 0.1	99.9 ± 0	
	G	0 ± 0	16 ± 2.8	0 ± 0	0 ± 0	
	T	0.2 ± 0	42.1 ± 5.3	1.3 ± 0.1	0 ± 0	

FIGURE 56

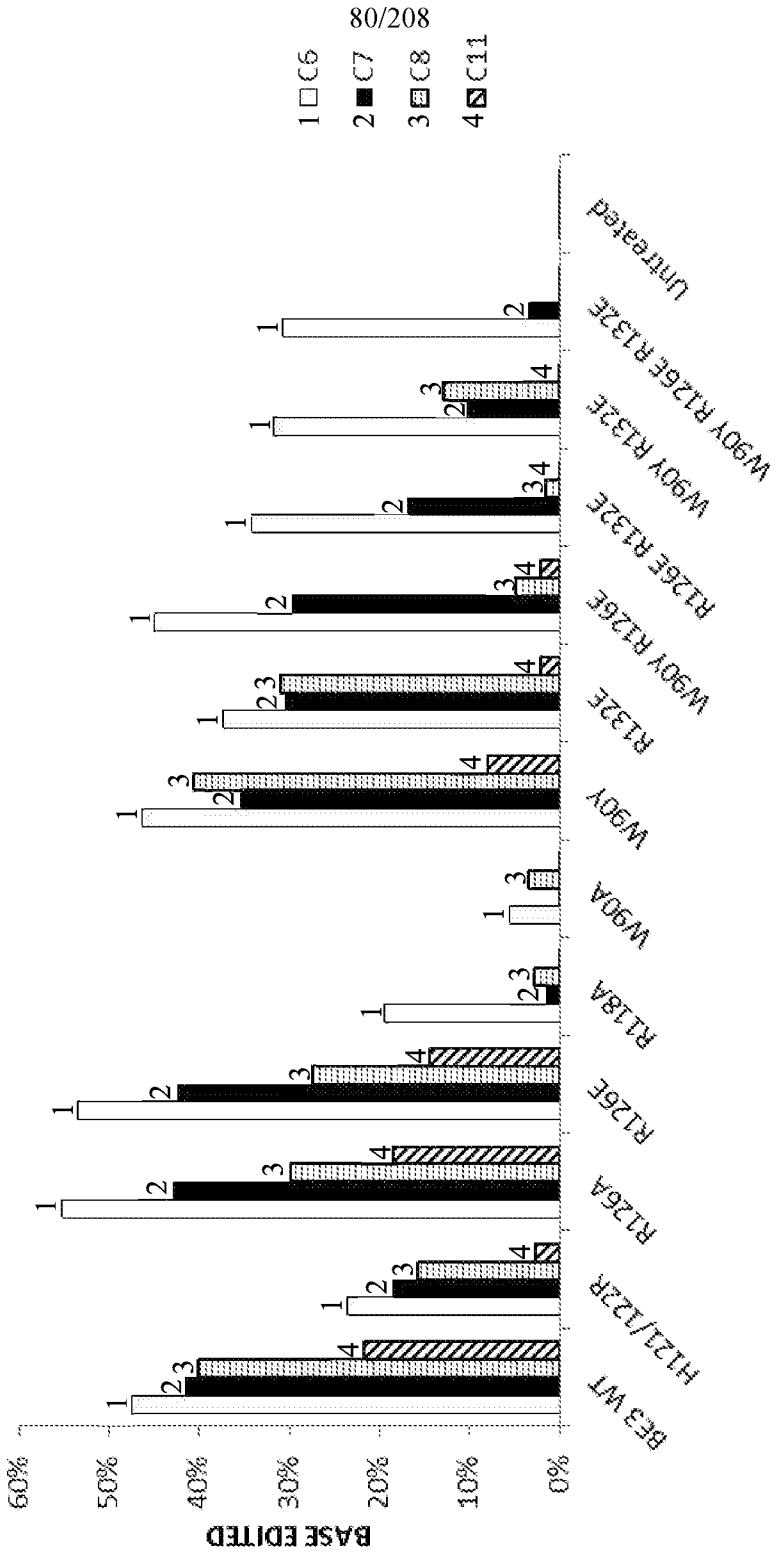

1 C6
2 C7
3 C8
4 C11
BASE EDITED
60%
50%
40%
30%
20%
10%
0%
BE3 WT
H121/122R
R126A
R126E
R18A
W90A
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
Untreated

FIGURE 57

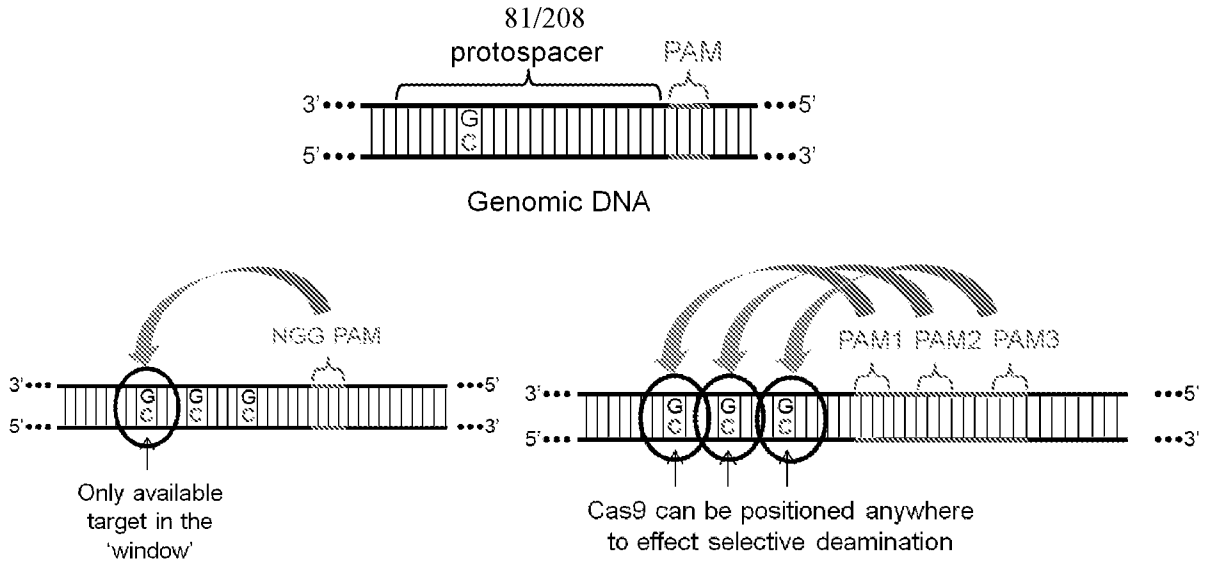


FIGURE 58

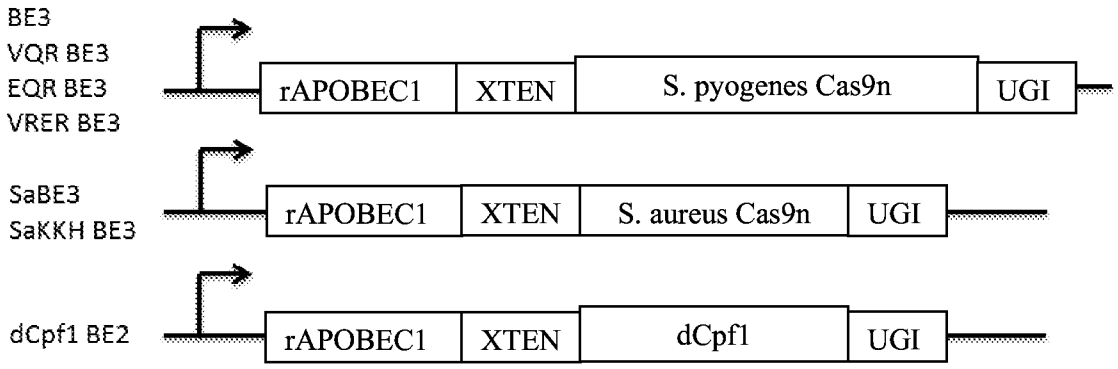


FIGURE 59

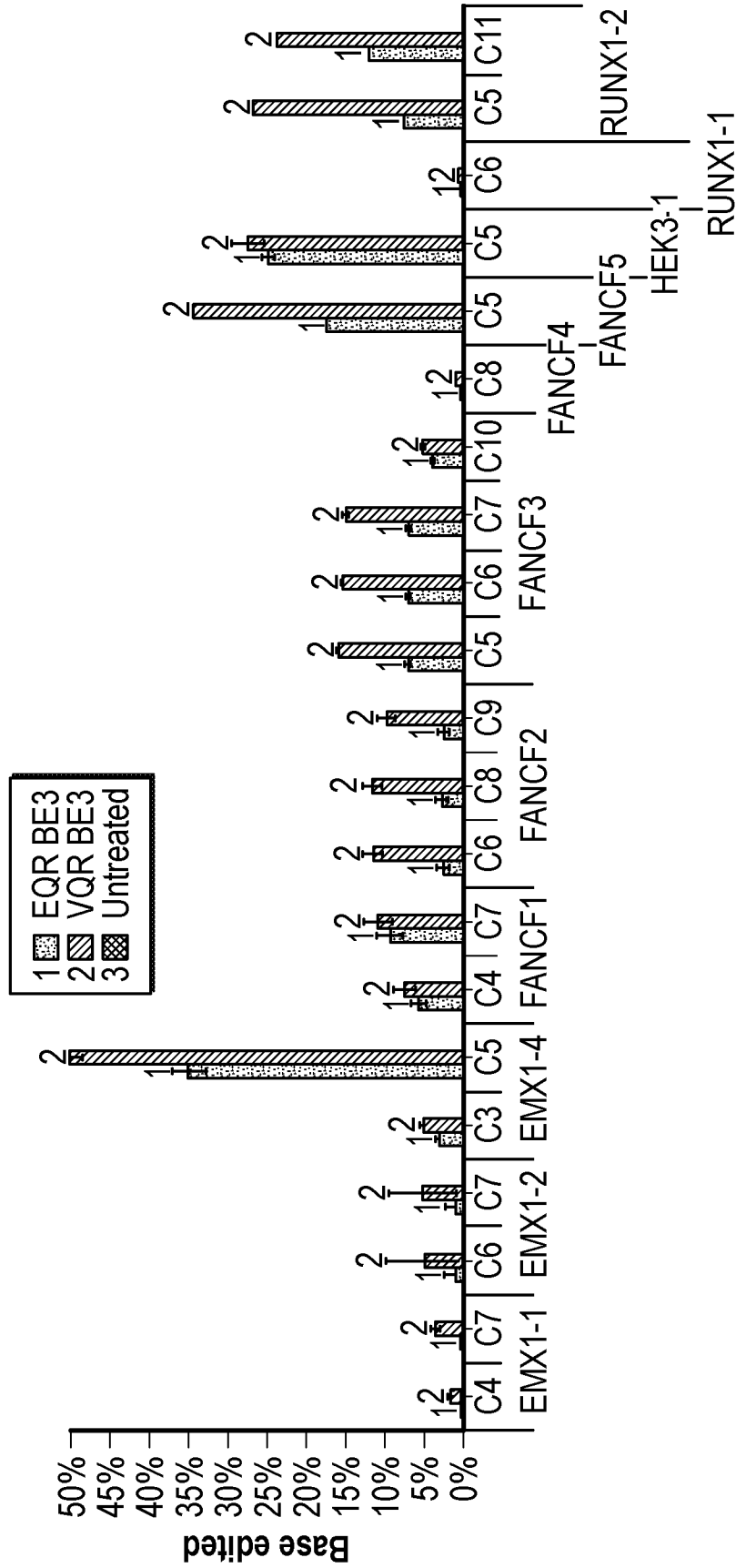

Base edited
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
1 EQR BE3
2 VQR BE3
3 Untreated
EMX1-1
EMX1-2
EMX1-4
FANCF1
FANCF2
FANCF3
FANCF4
FANCF5
HEK3-1
RUNX1-1
RUNX1-2
C4
C7
C6
C3
C5
C8
C9
C10
C11

FIGURE 60

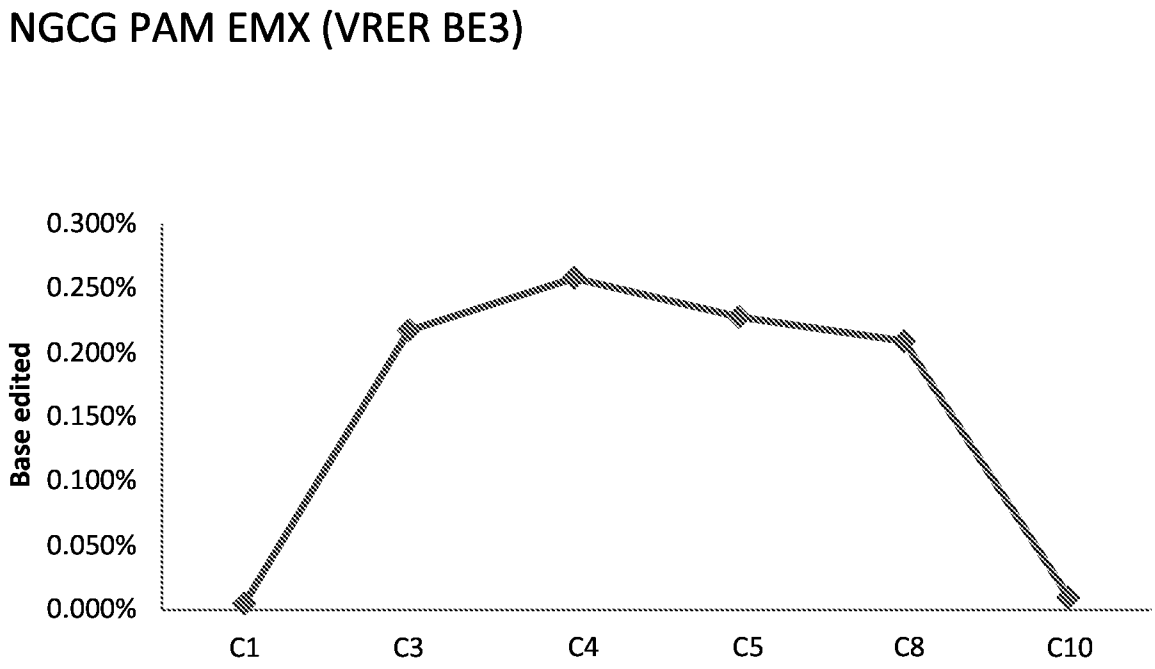

NGCG PAM EMX (VRER BE3)
Base edited
0.300%
0.250%
0.200%
0.150%
0.100%
0.050%
0.000%
C1
C3
C4
C5
C8
C10

FIGURE 61

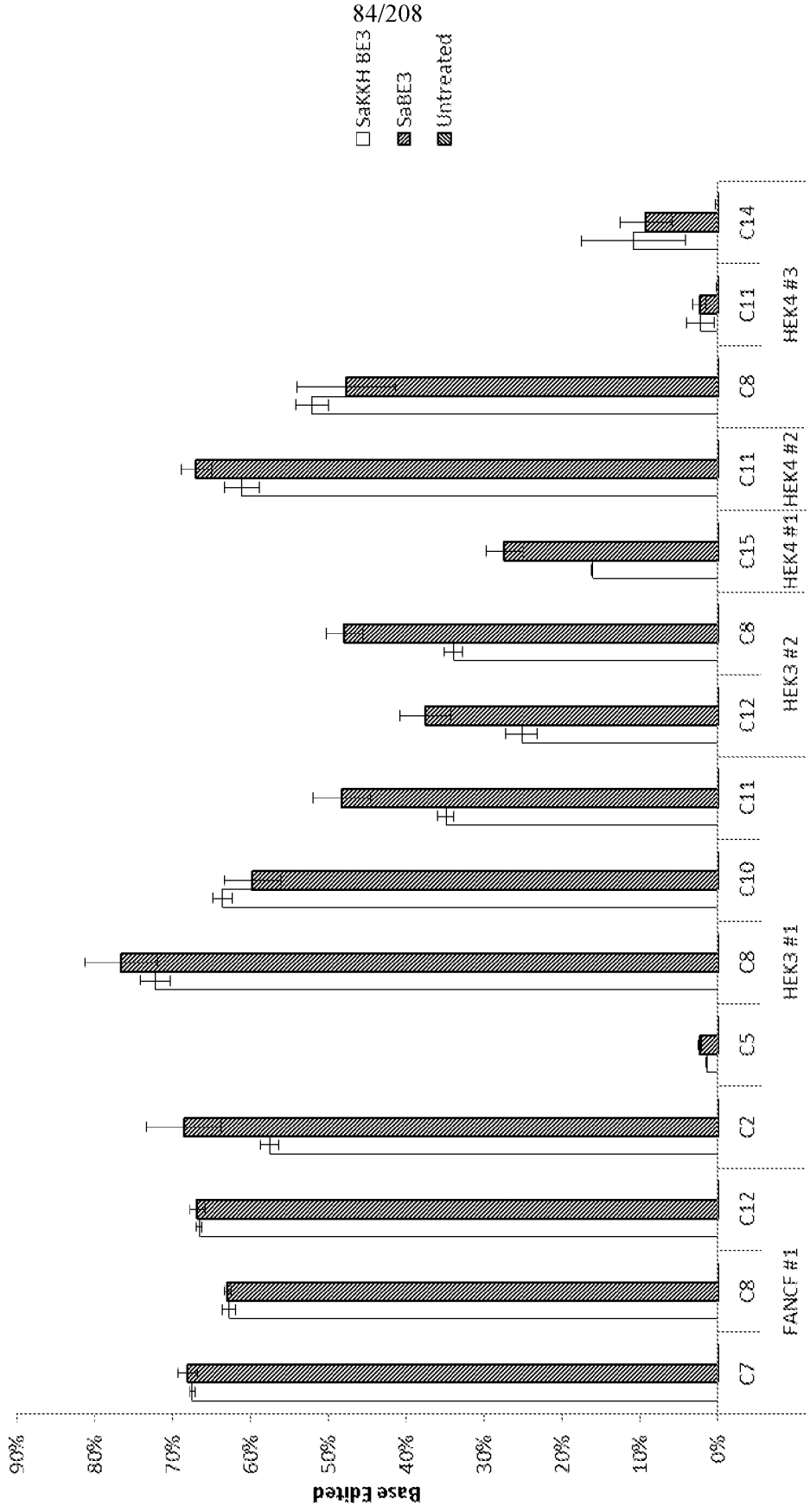

SaKKH BE3
SaBE3
Untreated
Base Edited
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
C7
C8
C12
C2
C5
C8
C10
C11
C12
C8
C15
C11
C8
C11
C14
FANCF #1
HEK3 #1
HEK3 #2
HEK4 #1
HEK4 #2
HEK4 #3

FIGURE 62

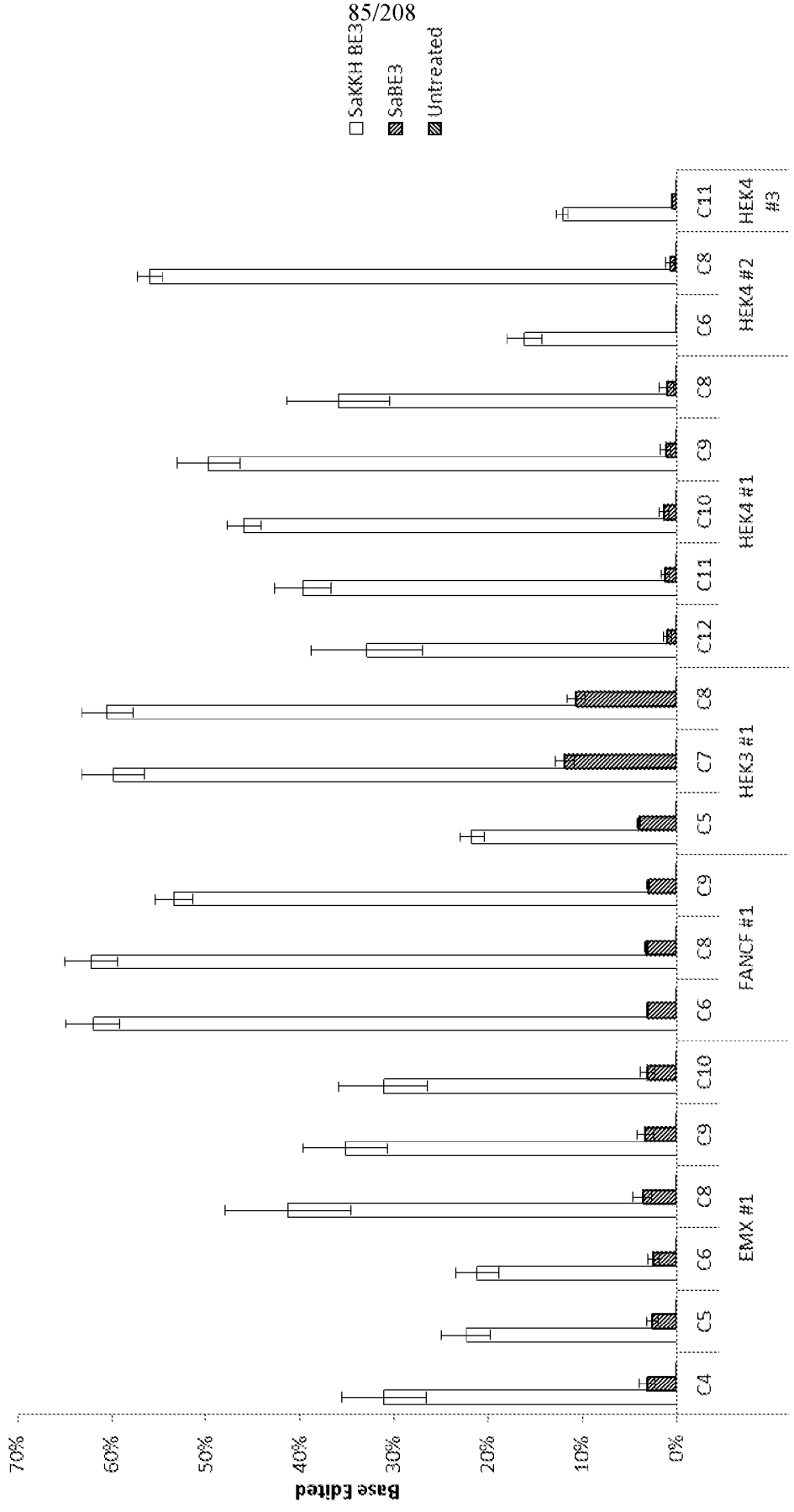

SaKKH BE3
SaBE3
Untreated
Base Edited
70%
60%
50%
40%
30%
20%
10%
0%
C4
C5
C6
C8
C9
C10
EMX #1
C6
C8
C9
FAMCF #1
C5
C7
C8
HEK3 #1
C12
C11
C10
C9
C8
HEK4 #1
C6
C8
HEK4 #2
C11
HEK4 #3

FIGURE 63

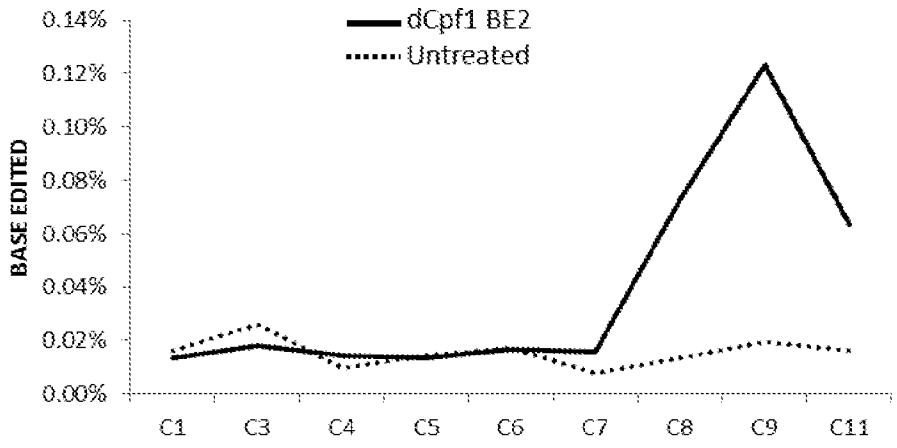


FIGURE 64A

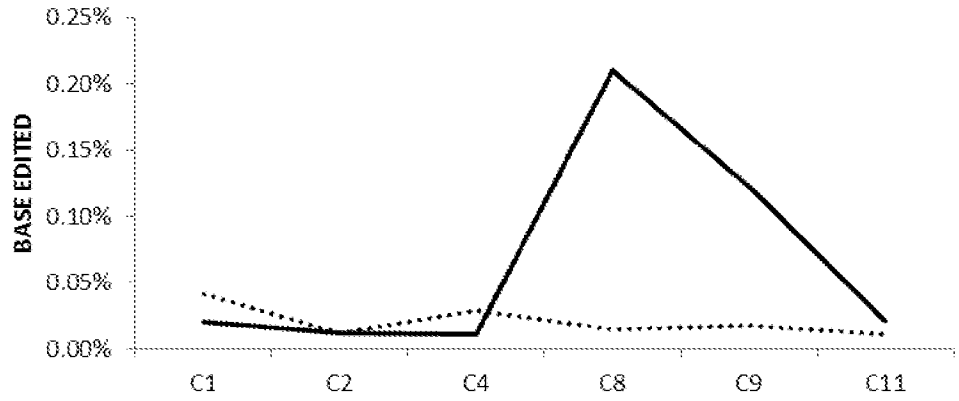


FIGURE 64B

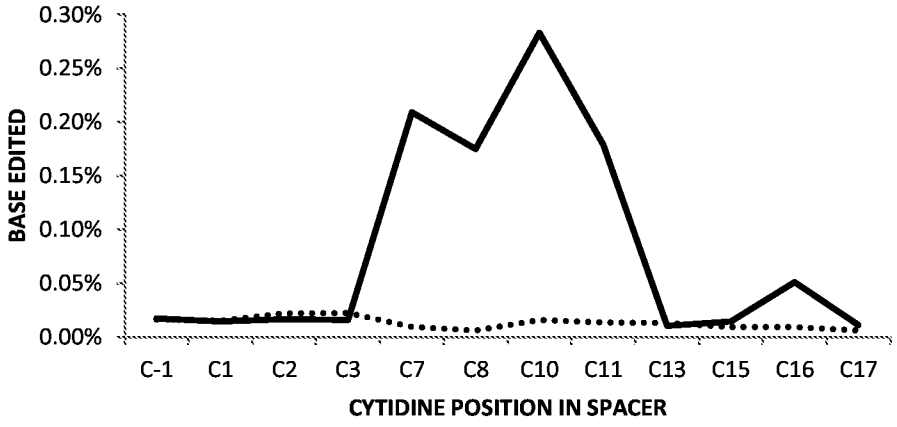


FIGURE 64C

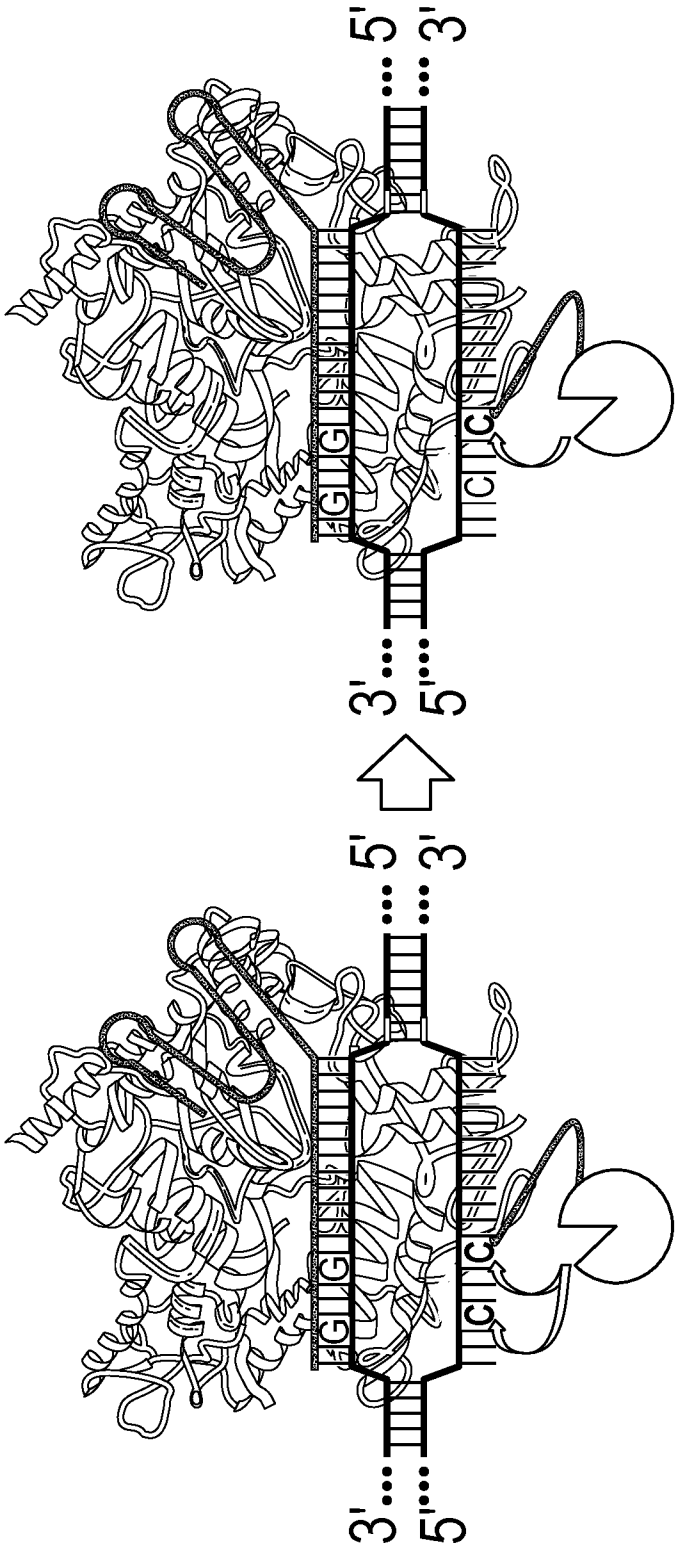


FIGURE 65

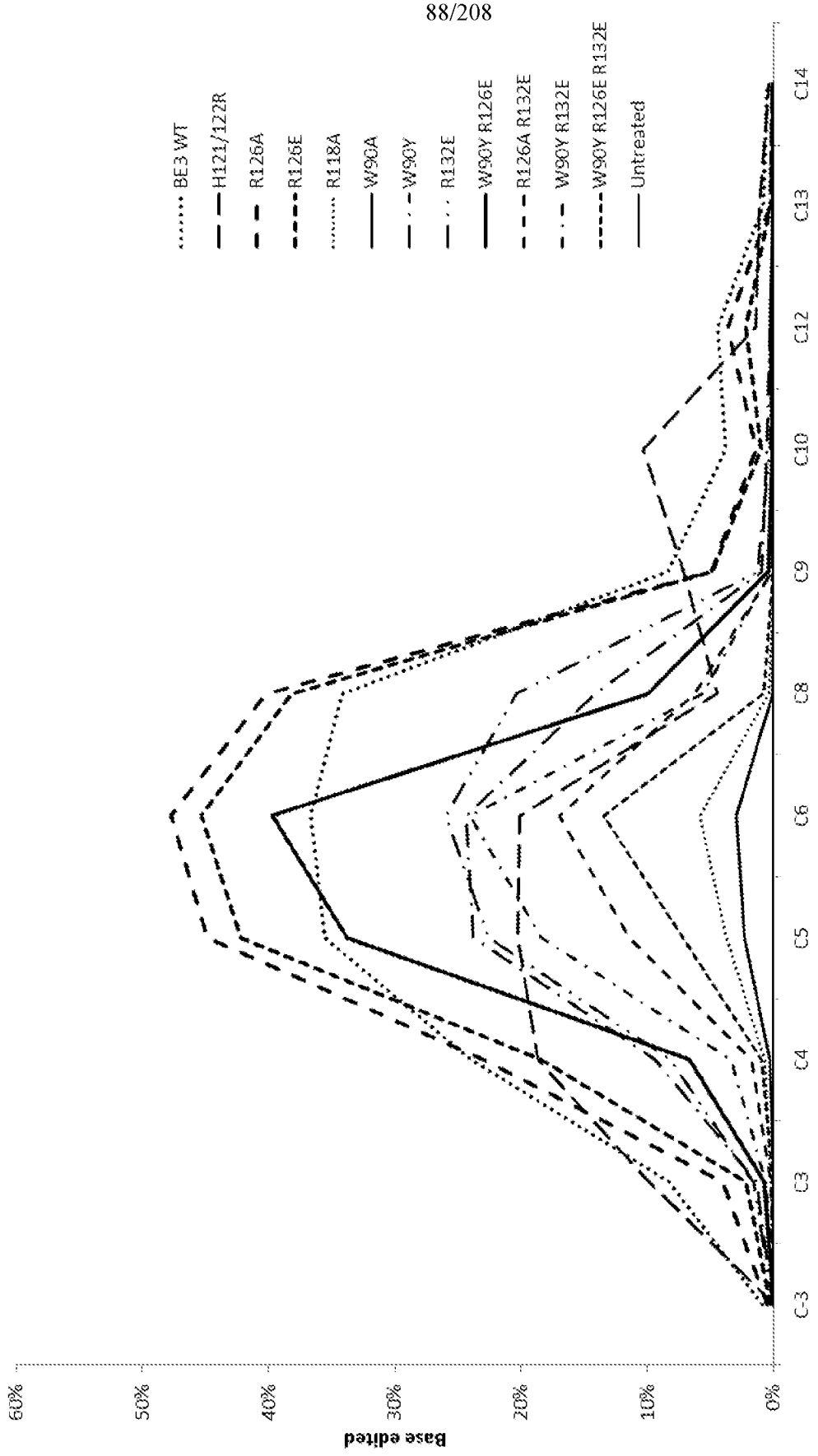

BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
R132E
W90Y R126E
R126A R132E
W90Y R132E
W90Y R126E R132E
Untreated
Base edited
60%
50%
40%
30%
20%
10%
0%
C-3
C3
C4
C5
C6
C8
C9
C10
C12
C13
C14

FIGURE 66

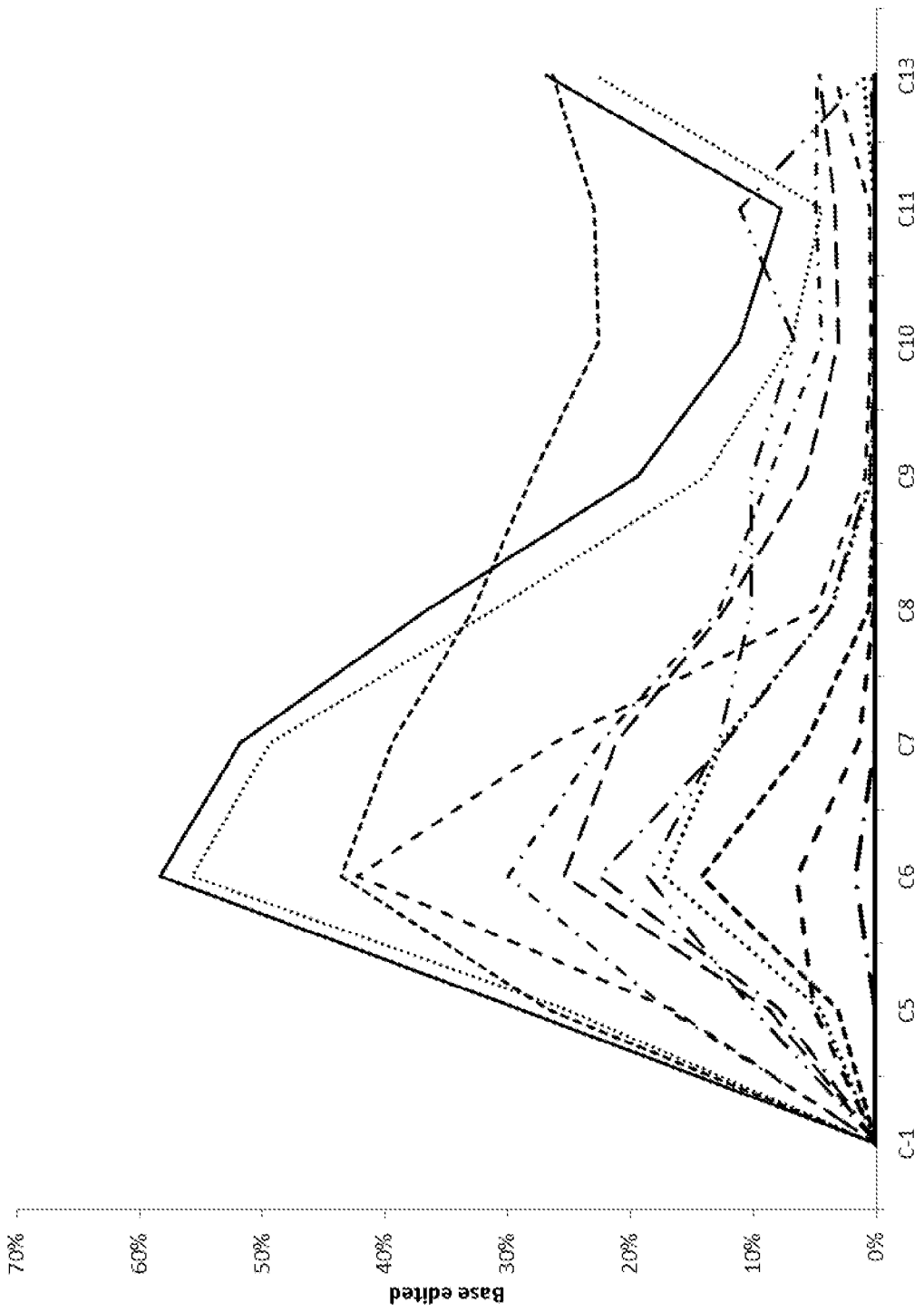

BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
R132E
W90Y R126E
R126A R132E
W90Y R132E
W90Y R126E R132E
Untreated
Base edited
70%
60%
50%
40%
30%
20%
10%
0%
C-1
C5
C6
C7
C8
C9
C10
C11
C13

FIGURE 67

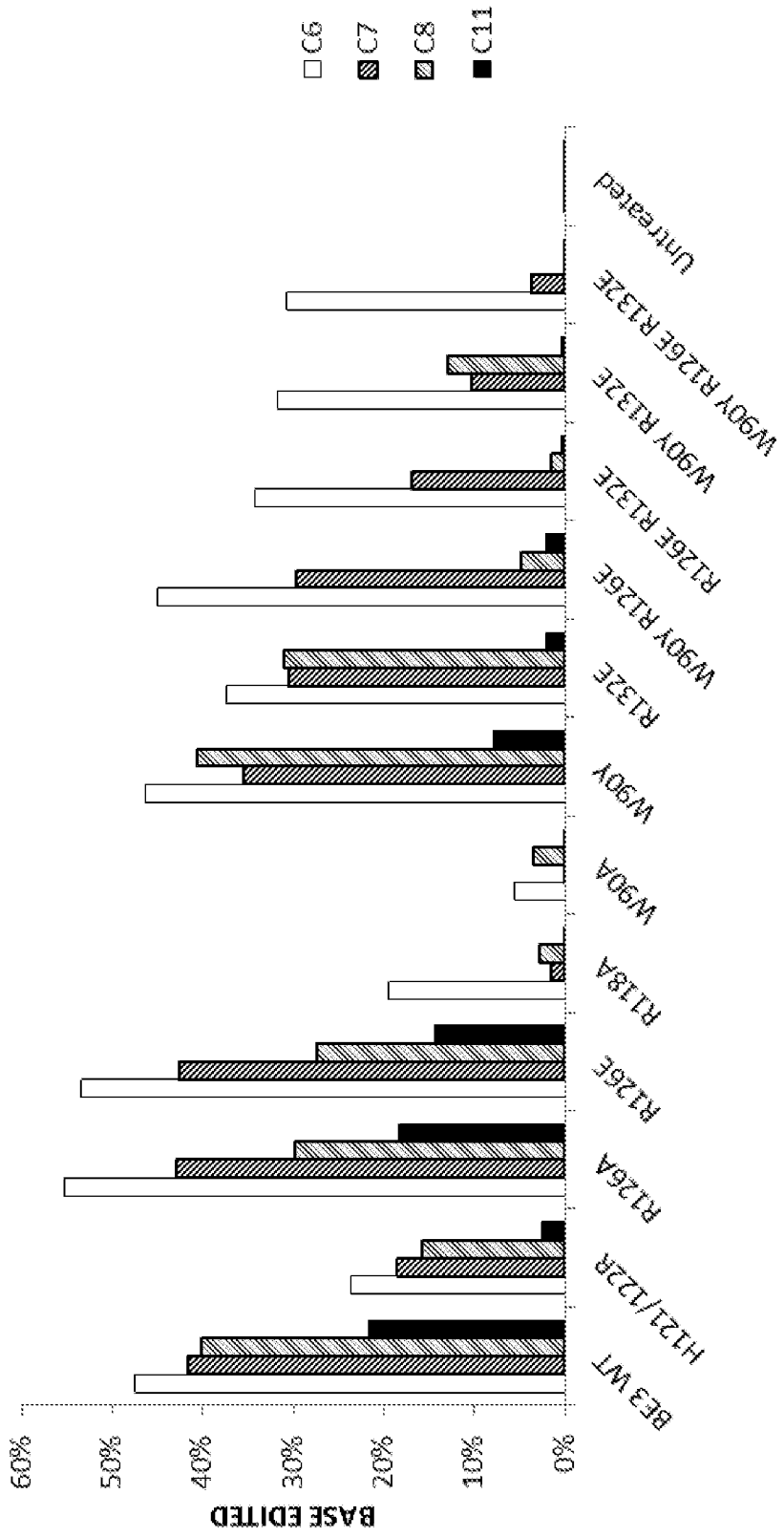

C6
C7
C8
C11
BASE EDITED
60%
50%
40%
30%
20%
10%
0%
BE3 WT
H121/122R
R126A
R126E
R18A
W90A
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
Untreated

FIGURE 68

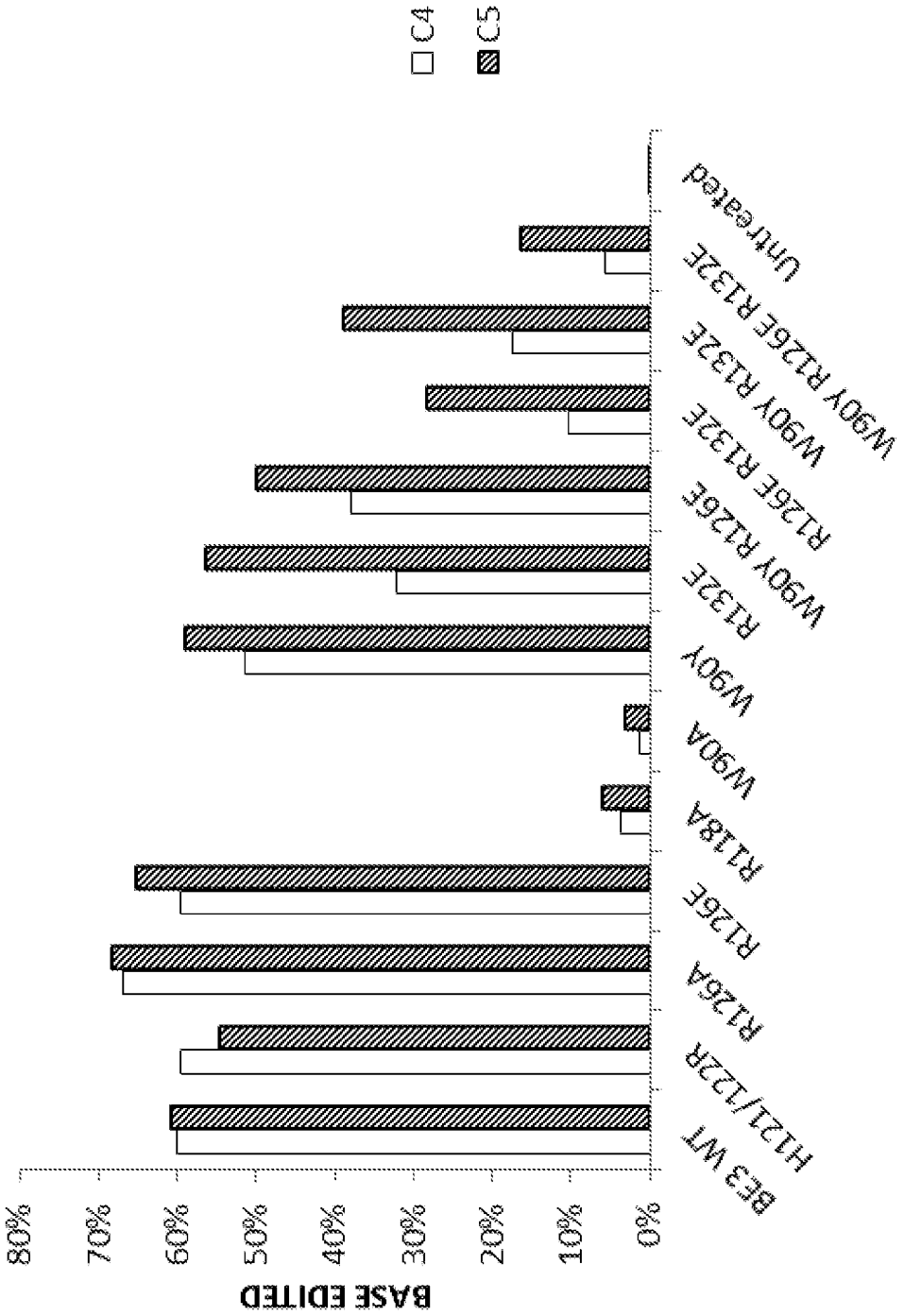

C4
C5
BASE EDITED
80%
70%
60%
50%
40%
30%
20%
10%
0%
BE3 WT
H121/122R
R126A
R126E
R18A
W90A
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
Untreated

FIGURE 69

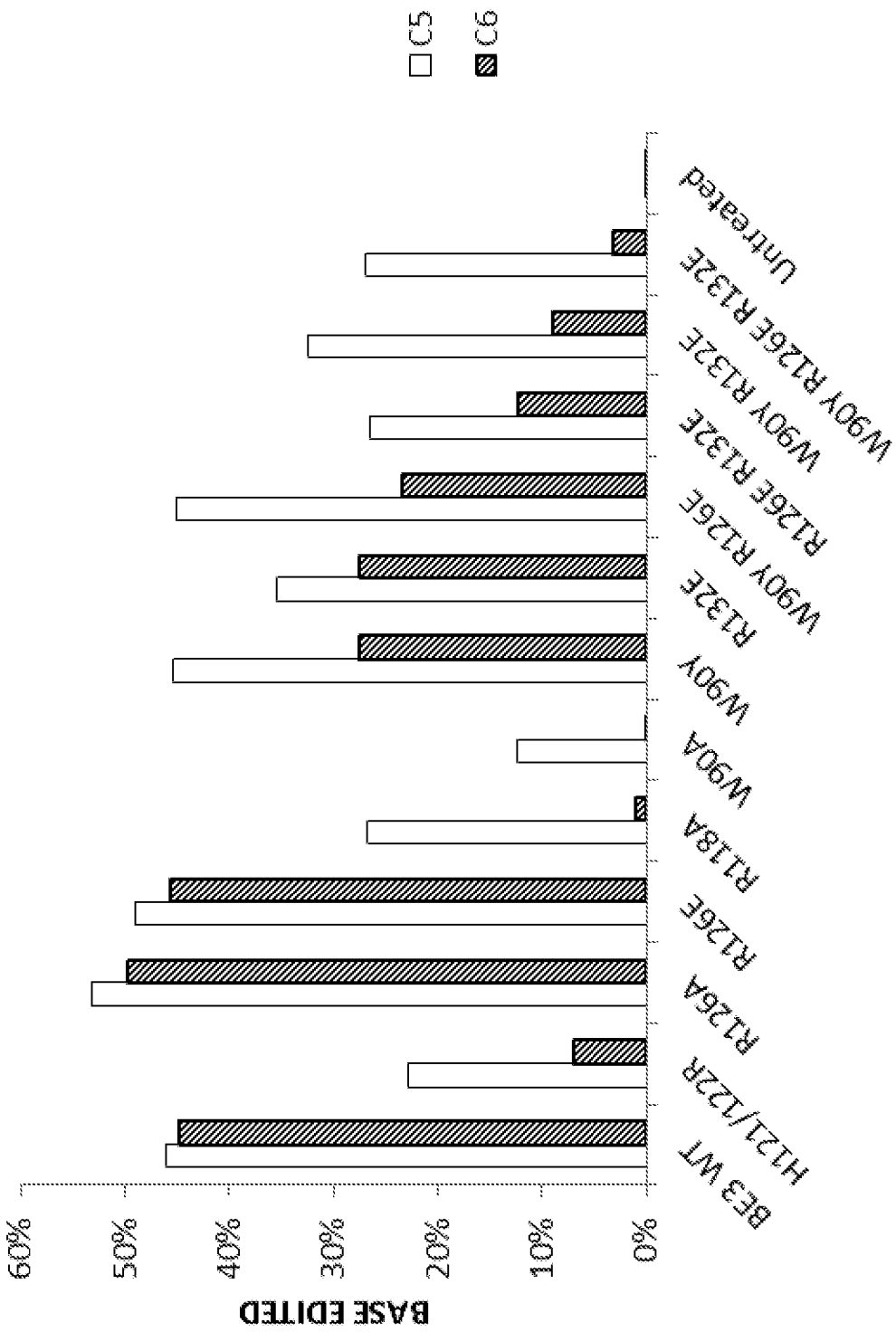

C5
C6
BASE EDITED
60%
50%
40%
30%
20%
10%
0%
BE3 WT
H122/122R
R126A
R126E
R18A
W90A
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
Untreated

FIGURE 70

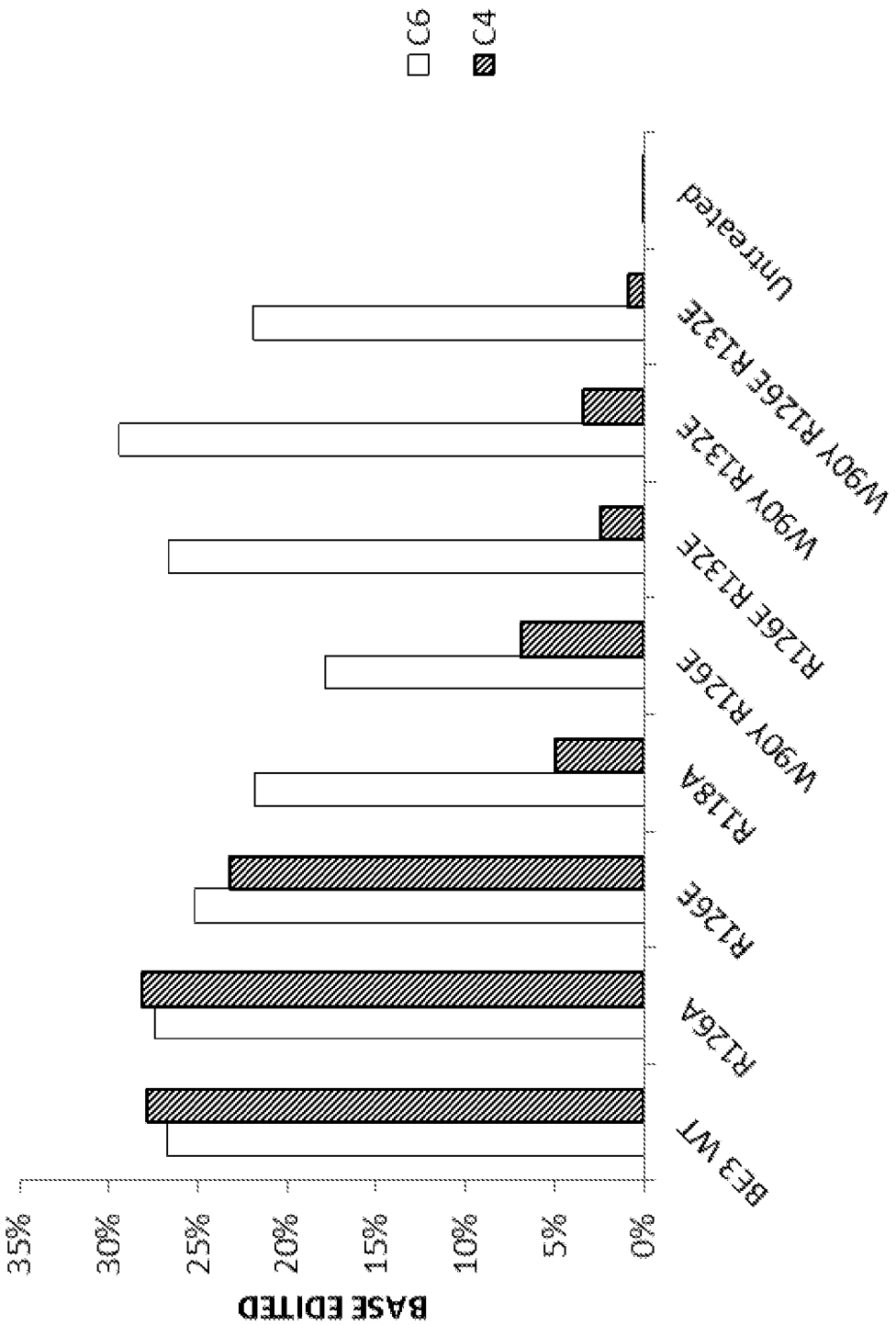

C6
C4
BASE EDITED
35%
30%
25%
20%
15%
10%
5%
0%
BE3 WT
R126A
R126E
R118A
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
Untreated

FIGURE 71

BE3

		Lys			Arg → Cys			Leu → Leu			Ala			Tyr			Gln			
Base	G	A	A	G	C	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C
A	0.2	99.9	99.8	0.0	1.0	0.1	1.8	1.2	0.0	0.0	0.1	0.1	99.8	0.0	0.0	0.1	0.1	99.9	0.1	0.0
C	0.0	0.0	0.1	0.0	38.1	0.0	49.8	52.3	0.0	0.0	0.0	99.7	0.1	0.0	0.1	0.0	0.0	0.0	99.9	99.9
G	99.8	0.0	0.0	99.9	1.8	99.8	1.3	0.7	0.1	99.9	99.8	0.0	0.1	99.9	0.1	99.9	0.1	0.0	0.0	0.0
T	0.0	0.0	0.1	0.1	59.2	0.1	47.0	45.8	99.9	0.1	0.1	0.2	0.0	0.0	99.8	0.0	99.8	0.0	0.1	0.1

BE3 W90Y R132E

		Lys			Arg → Cys			Leu → Leu			Ala			Tyr			Gln			
Base	G	A	A	G	C	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C
A	0.0	99.9	99.9	0.0	0.5	0.1	0.3	0.1	0.0	0.0	0.0	0.1	99.8	0.0	0.0	0.0	0.1	99.9	0.1	0.0
C	0.0	0.0	0.0	0.1	78.0	0.0	94.8	98.9	0.0	0.0	0.0	99.8	0.1	0.0	0.0	0.0	0.0	0.0	99.9	99.9
G	100.0	0.0	0.0	99.9	0.5	99.9	0.1	0.1	0.0	99.9	99.9	0.0	0.1	100.0	0.1	99.9	0.0	0.0	0.0	0.0
T	0.0	0.0	0.0	0.0	21.0	0.0	4.8	0.9	99.9	0.0	0.0	0.1	0.0	0.0	99.8	0.0	99.9	0.0	0.0	0.0

FIGURE 72

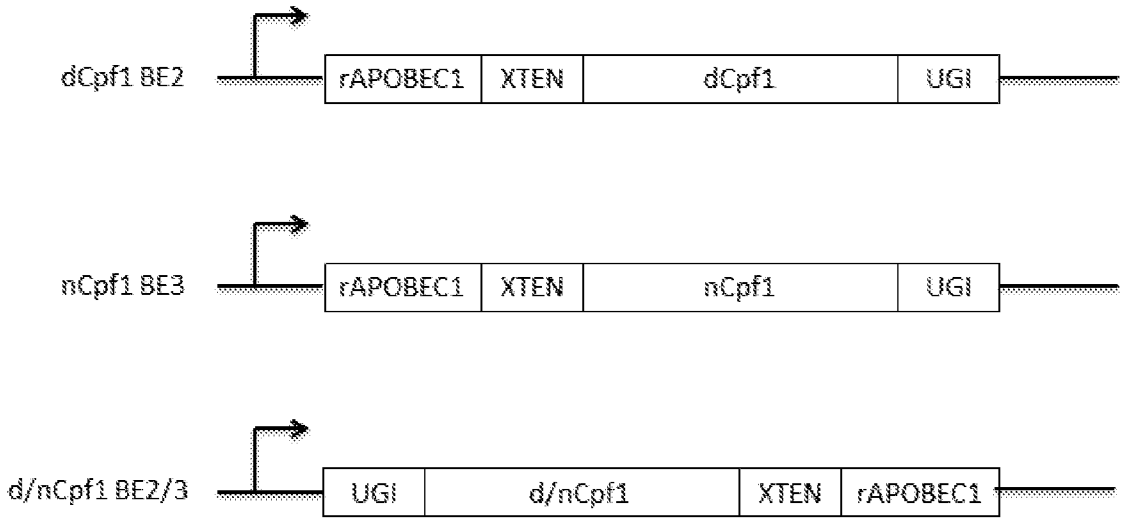
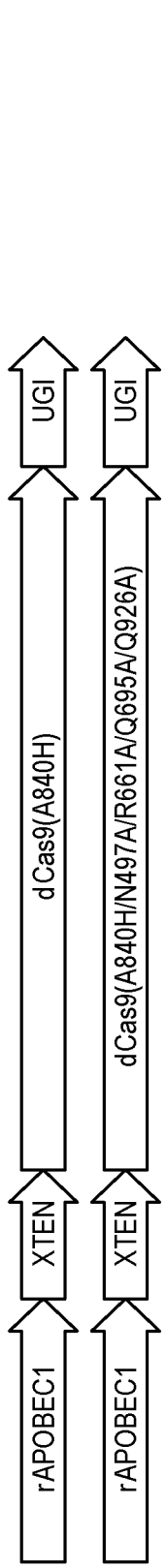

dCpf1 BE2
rAPOBEC1
XTEN
dCpf1
UGI
nCpf1 BE3
rAPOBEC1
XTEN
nCpf1
UGI
d/nCpf1 BE2/3
UGI
d/nCpf1
XTEN
rAPOBEC1

FIGURE 73



EMX1		C_3	C_6	C_{12}	indel%
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9±0	99.9±0	99.9±0	
	G	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	

BE3		C_3	C_6	C_{12}	indel%
	A	1.3 ± 0.1	0.7 ± 0.2	0 ± 0	2.6±0.3
	C	51.7 ± 1	55.2±0.1	99.9 ± 0	
	G	4 ± 0.7	1.9 ± 0.4	0 ± 0	
	T	42.8 ± 0.2	41.9 ± 0.4	0 ± 0	

HF BE3		C_3	C_6	C_{12}	indel%
	A	1.5 ± 0.7	0.5 ± 0.1	0 ± 0	1 ± 0.1
	C	44.8 ± 8.5	49.9 ± 6.6	99.9 ± 0	
	G	4.2 ± 1.2	0.6 ± 0.2	0 ± 0	
	T	49.3 ± 6.5	48.7 ± 6.3	0 ± 0	

FANCF		C_6	C_7	C_8	C_{11}	indel%
untreated	A	0±0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9±0	99.9 ± 0	99.9±0	99.9 ± 0	
	G	0±0	0 ± 0	0 ± 0	0 ± 0	
	T	0±0	0 ± 0	0 ± 0	0 ± 0	

BE3		C_6	C_7	C_8	C_{11}	5.8 ± 0.9
	A	0.8 ± 0.1	1.2 ± 0.3	1.2 ± 0.1	0.2 ± 0.1	
	C	64.7 ± 5.9	67.9 ± 5.1	68.9 ± 5.2	85.4 ± 2.7	
	G	0.6 ± 0	0.7 ± 0	0.7 ± 0.1	0.1 ± 0	
	T	33.8 ± 5.7	29.9 ± 4.8	29 ± 5	14.1 ± 2.6	

HF BE3		C_6	C_7	C_8	C_{11}	5.9 ± 0.7
	A	0.9 ± 0.2	2 ± 0.4	0.9 ± 0.2	0.4 ± 0	
	C	52.5 ± 8.4	57.7 ± 7.6	61.9 ± 6.8	84.7 ± 2.4	
	G	1.4 ± 0.2	0.6 ± 0	0.5 ± 0.1	0.3 ± 0	
	T	45 ± 8	39.5 ± 7.2	36.5 ± 6.5	14.4 ± 2.2	

RNF2		C_3	C_6	C_{12}	indel%
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	

BE3		C_3	C_6	C_{12}	2.3 ± 0.3
	A	0.4 ± 0	1.5 ± 0.2	0.1 ± 0	
	C	70.9 ± 2.7	45 ± 0	95 ± 0.2	
	G	0 ± 0	12.6 ± 2.4	0.1 ± 0	
	T	28.5 ± 2.7	40.7 ± 2.7	4.6 ± 0.2	

HF BE3		C_3	C_6	C_{12}	0.5 ± 0.1
	A	0 ± 0	0.2 ± 0	0 ± 0	
	C	76.8 ± 1.8	64.5 ± 3.3	95.4 ± 0.1	
	G	0 ± 0	2.4 ± 0.5	0 ± 0	
	T	23 ± 1.8	32.7 ± 2.6	4.5 ± 0.1	

numbers are μ ± σ from three independent replicates

FIGURE 74

HEK3		C_6	C_7	C_8	C_{11}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	

HEK4		C_6	C_7	C_8	C_{11}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	

BE3						
	A	0 ± 0	1.6 ± 0.2	1.5 ± 0.3	0.2 ± 0	2.7 ± 0.4
	C	98.7 ± 0.2	48.6 ± 5.4	40.8 ± 6.7	98.6 ± 0.1	
	G	0 ± 0	1.4 ± 0.2	10.5 ± 1.2	0.3 ± 0	
	T	1.1 ± 0.2	48.2 ± 4.9	47 ± 5.2	0.7 ± 0.1	

BE3						
	A	0 ± 0	5.4 ± 1.3	0 ± 0	0 ± 0	3 ± 0.7
	C	98.7 ± 0	44.5 ± 8.2	98.9 ± 0	99.8 ± 0	
	G	0 ± 0	18.1 ± 2.3	0 ± 0	0 ± 0	
	T	1.1 ± 0.1	31.9 ± 4.8	0.9 ± 0.1	0.1 ± 0	

HF BE3						
	A	0 ± 0	0.8 ± 0.2	1.1 ± 0.3	0.4 ± 0	4.1 ± 0.7
	C	97.8 ± 0.4	65 ± 3.8	40.4 ± 6.3	95.9 ± 0.5	
	G	0 ± 0	0.9 ± 0.2	6.1 ± 1.1	0.7 ± 0.2	
	T	2 ± 0.4	33 ± 3.4	52.2 ± 5.2	2.7 ± 0.2	

HF BE3						
	A	0 ± 0	5.2 ± 0.8	0 ± 0	0 ± 0	1.3 ± 0
	C	99.6 ± 0	36.6 ± 9	98.5 ± 0.1	99.9 ± 0	
	G	0 ± 0	16 ± 2.8	0 ± 0	0 ± 0	
	T	0.2 ± 0	42.1 ± 5.3	1.3 ± 0.1	0 ± 0	

numbers are μ ± σ from three independent replicates

FIGURE 75

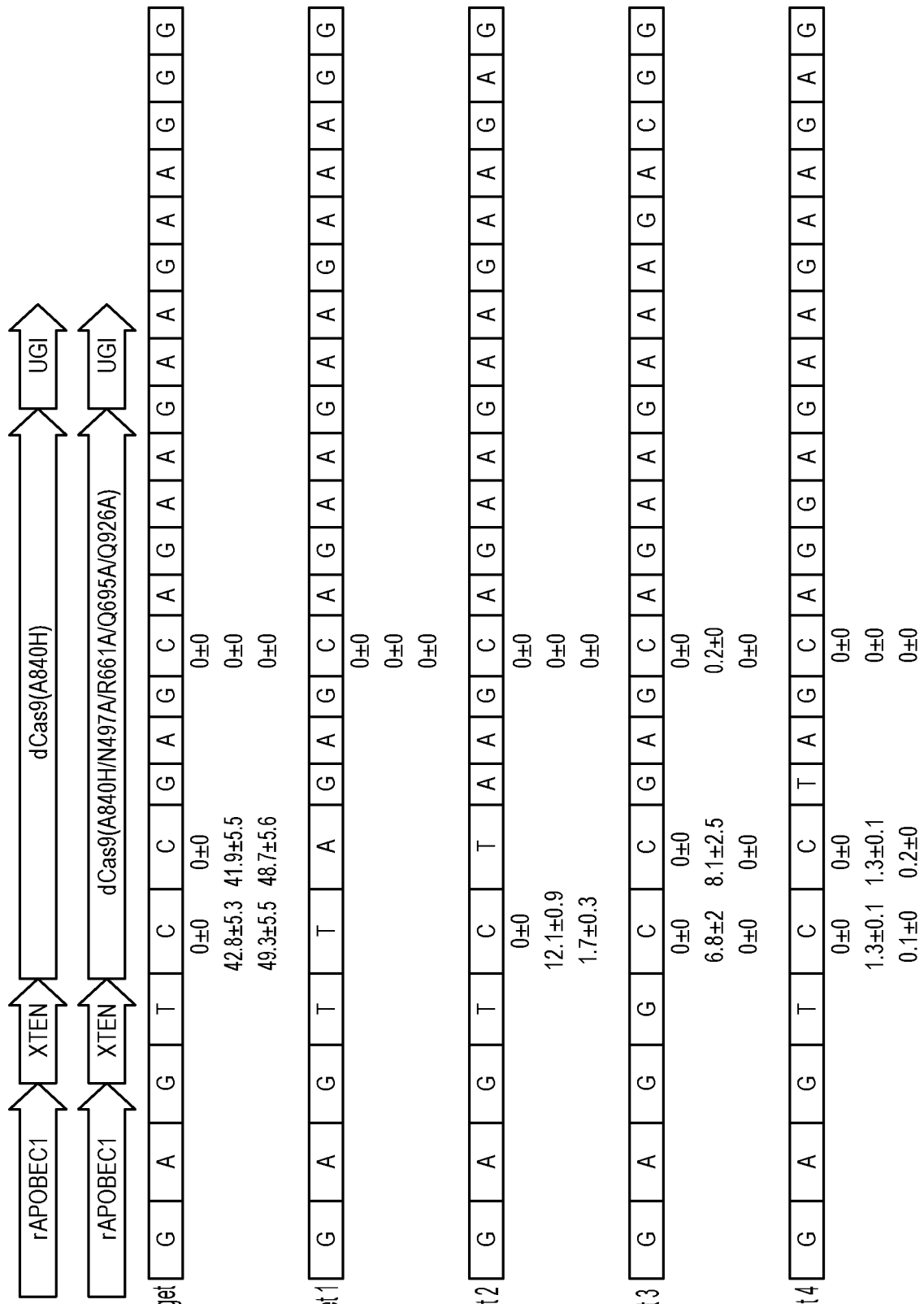

BE3
rAPOBEC1
XTEN
dCas9(A840H)
UGI
HF-BE3
rAPOBEC1
XTEN
dCas9(A840H/N497A/R661A/Q695A/Q926A)
UGI
EMX1 on target
untreated
BE3
HF-BE3
EMX1 off target 1
untreated
BE3
HF-BE3
EMX1 off target 2
untreated
BE3
HF-BE3
EMX1 off target 3
untreated
BE3
HF-BE3
EMX1 off target 4
untreated
BE3
HF-BE3
G A G T C C G A G C A G A A G G G
0±0 0±0 0±0
42.8±5.3 41.9±5.5 0±0
49.3±5.5 48.7±5.6 0±0
G A G T T A G A G C A G A A G G G
0±0
0±0
0±0
G A G T C T A A G C A G A A G A G
0±0
12.1±0.9 0±0
1.7±0.3 0±0
G A G G C C G A G C A G A A G C G
0±0 0±0 0±0
6.8±2 8.1±2.5 0.2±0
0±0 0±0 0±0
G A G T C C T A G C A G G A A G A G
0±0 0±0 0±0
1.3±0.1 1.3±0.1 0±0
0.1±0 0.2±0 0±0

FIGURE 76

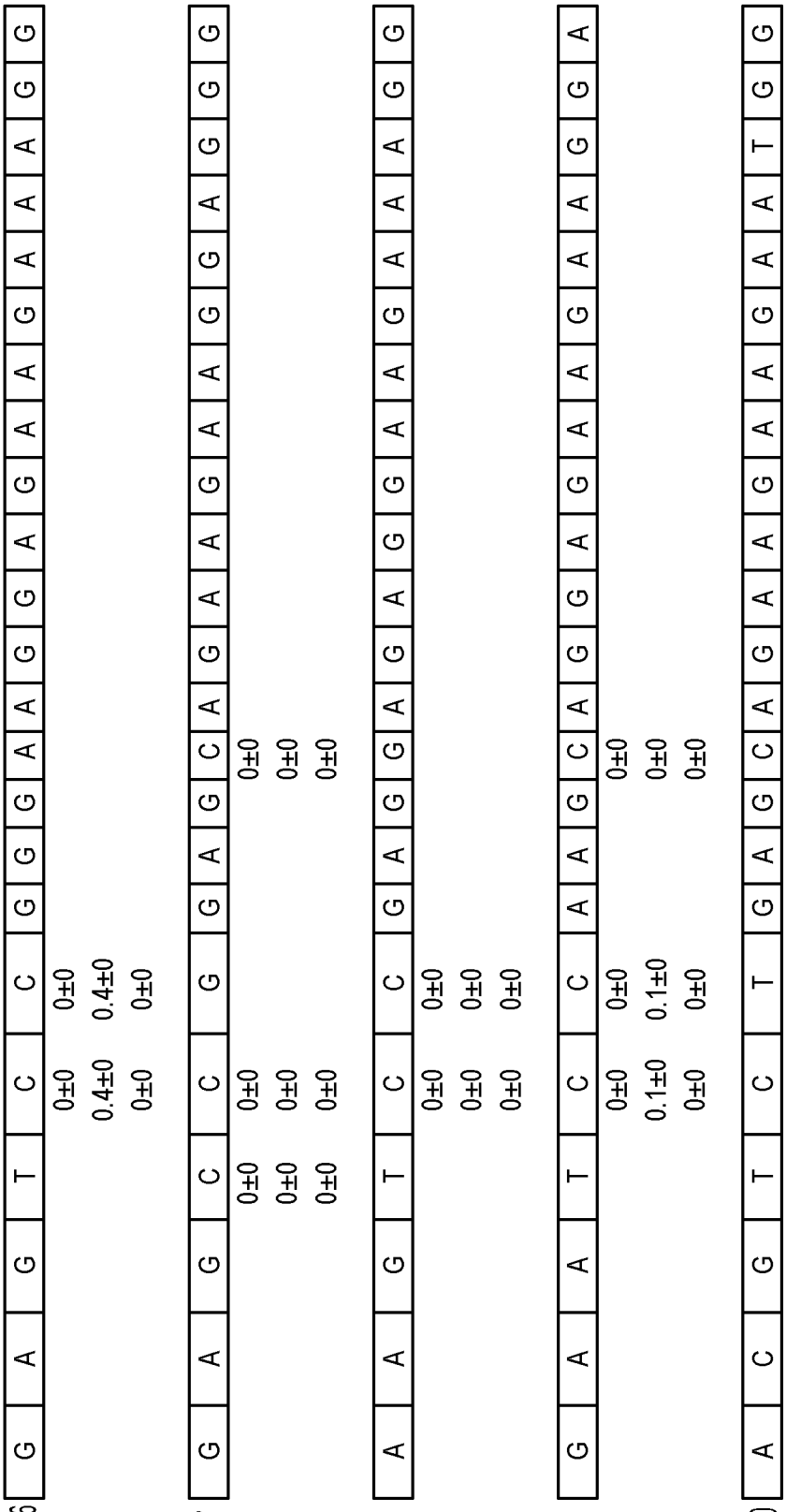


FIGURE 76 (CONTINUED)

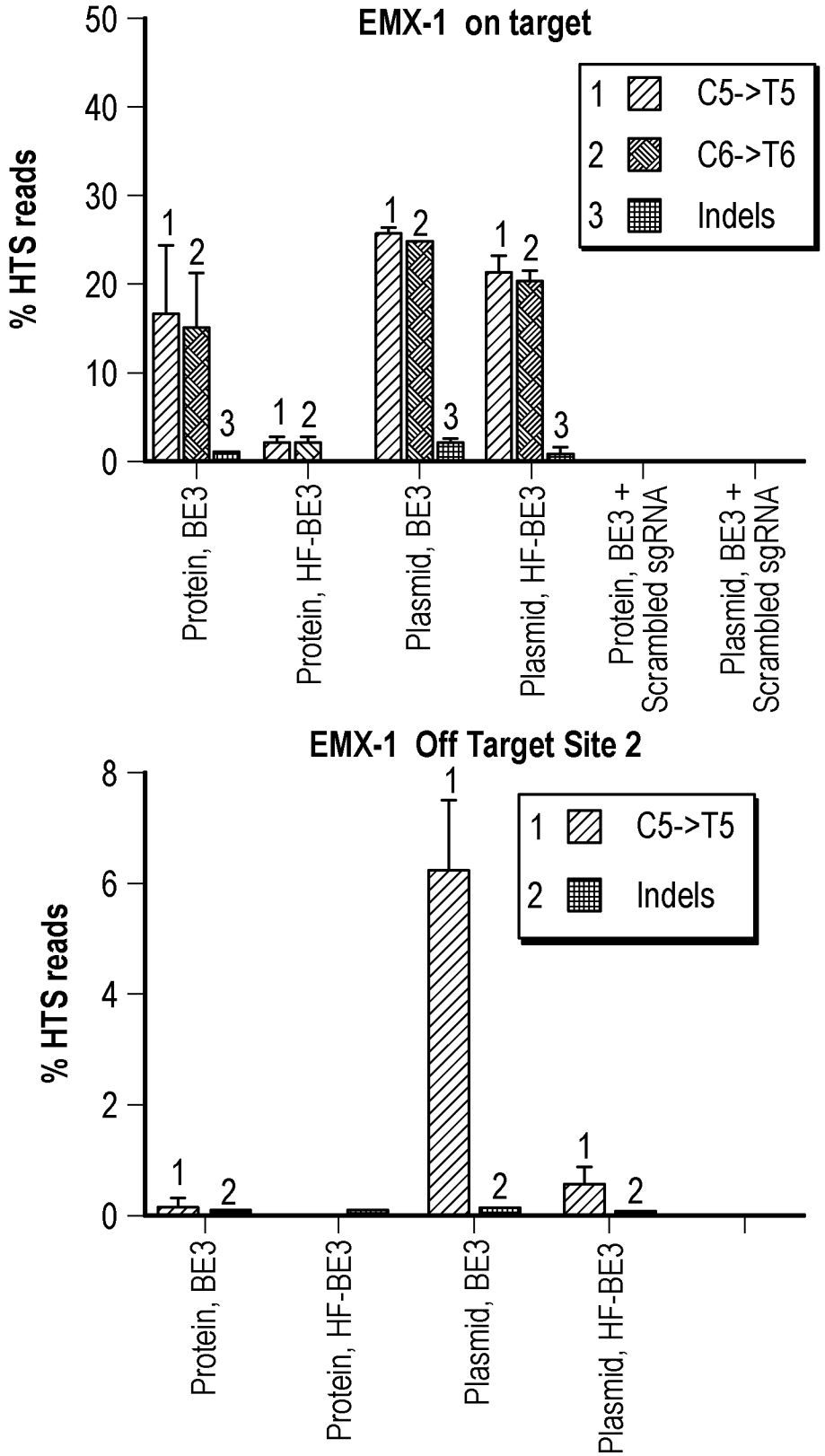

EMX-1 on target
% HTS reads
50
40
30
20
10
0
1 C5->T5
2 C6->T6
3 Indels
Protein, BE3
Protein, HF-BE3
Plasmid, BE3
Plasmid, HF-BE3
Protein, BE3 + Scrambled sgRNA
Plasmid, BE3 + Scrambled sgRNA
EMX-1 Off Target Site 2
% HTS reads
8
6
4
2
0
1 C5->T5
2 Indels
Protein, BE3
Protein, HF-BE3
Plasmid, BE3
Plasmid, HF-BE3

FIGURE 77

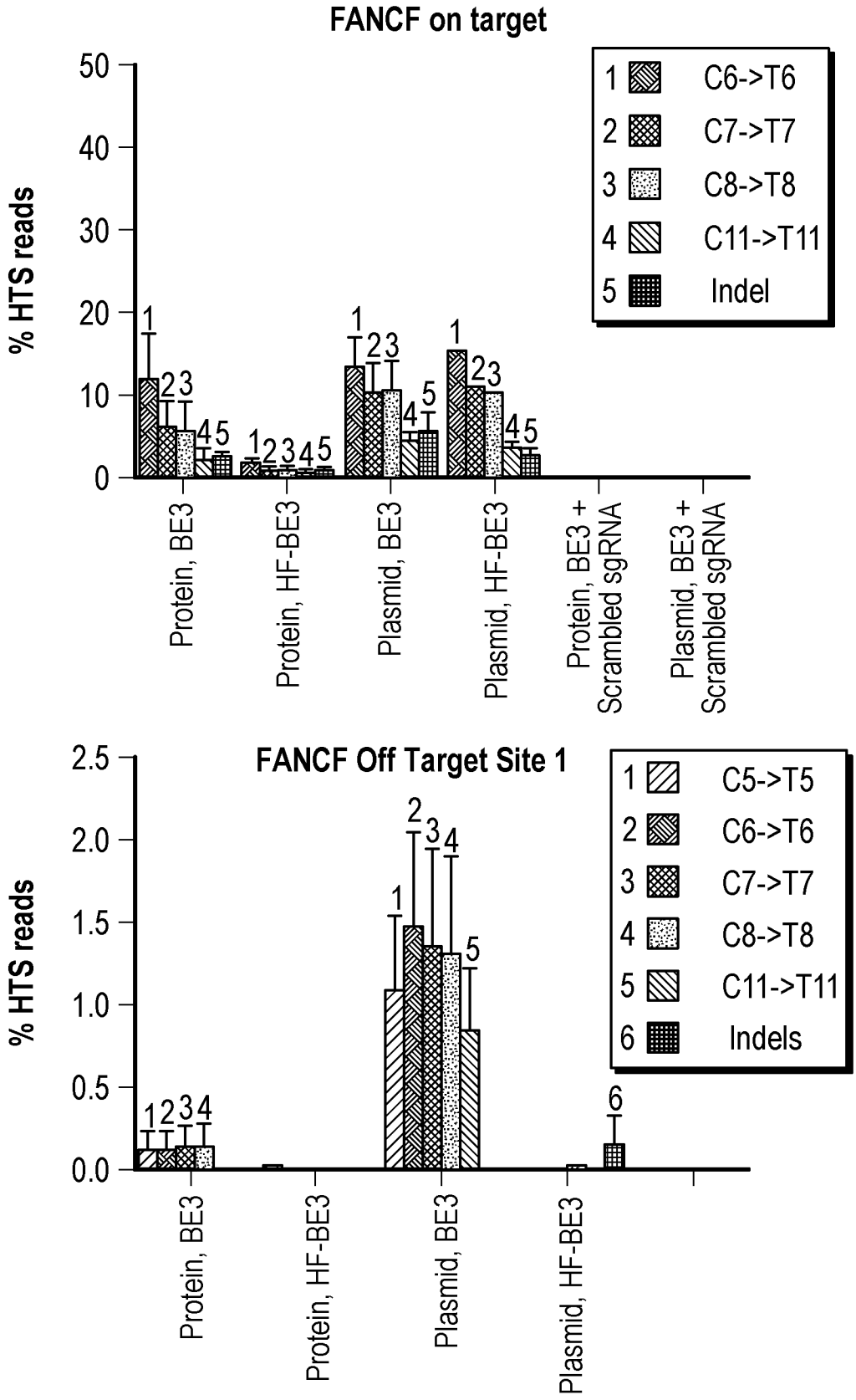

FANCF on target
% HTS reads
50
40
30
20
10
0
1 C6->T6
2 C7->T7
3 C8->T8
4 C11->T11
5 Indel
Protein, BE3
Protein, HF-BE3
Plasmid, BE3
Plasmid, HF-BE3
Protein, BE3 + Scrambled sgRNA
Plasmid, BE3 + Scrambled sgRNA
FANCF Off Target Site 1
% HTS reads
2.5
2.0
1.5
1.0
0.5
0.0
1 C5->T5
2 C6->T6
3 C7->T7
4 C8->T8
5 C11->T11
6 Indels
Protein, BE3
Protein, HF-BE3
Plasmid, BE3
Plasmid, HF-BE3

FIGURE 78

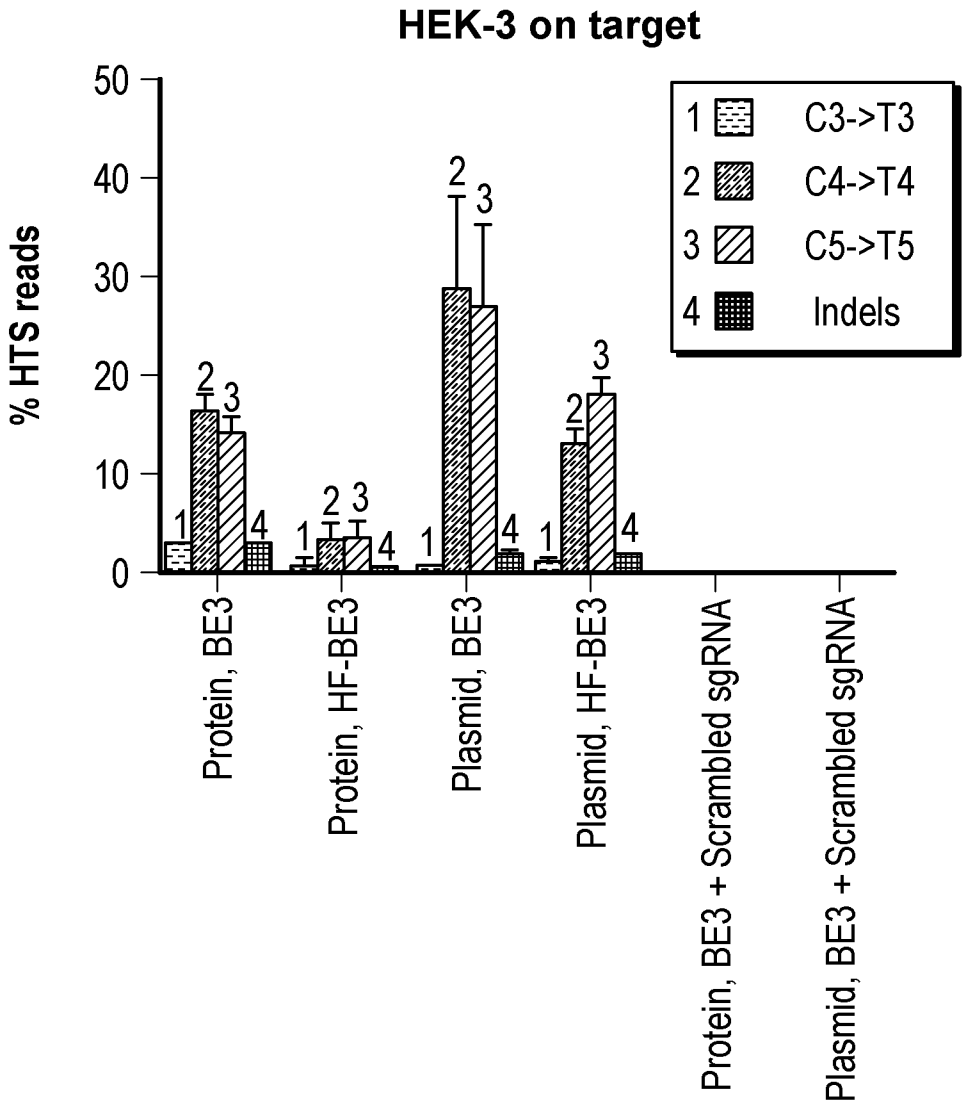

HEK-3 on target
% HTS reads
50
40
30
20
10
0
1 C3->T3
2 C4->T4
3 C5->T5
4 Indels
Protein, BE3
Protein, HF-BE3
Plasmid, BE3
Plasmid, HF-BE3
Protein, BE3 + Scrambled sgRNA
Plasmid, BE3 + Scrambled sgRNA

FIGURE 79

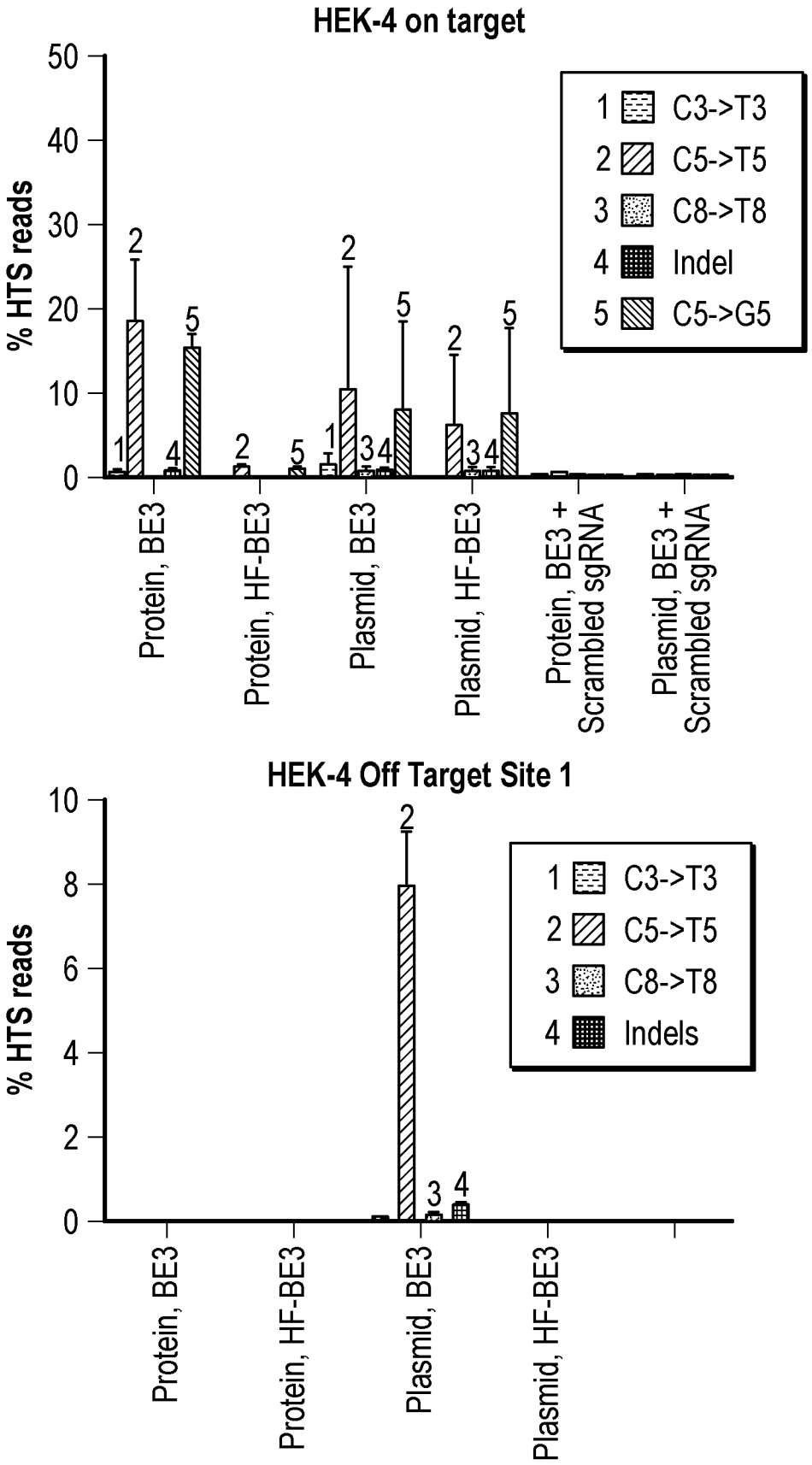

HEK-4 on target
% HTS reads
50
40
30
20
10
0
1 C3->T3
2 C5->T5
3 C8->T8
4 Indel
5 C5->G5
Protein, BE3
Protein, HF-BE3
Plasmid, BE3
Plasmid, HF-BE3
Protein, BE3 + Scrambled sgRNA
Plasmid, BE3 + Scrambled sgRNA
HEK-4 Off Target Site 1
% HTS reads
10
8
6
4
2
0
1 C3->T3
2 C5->T5
3 C8->T8
4 Indels
Protein, BE3
Protein, HF-BE3
Plasmid, BE3
Plasmid, HF-BE3

FIGURE 80

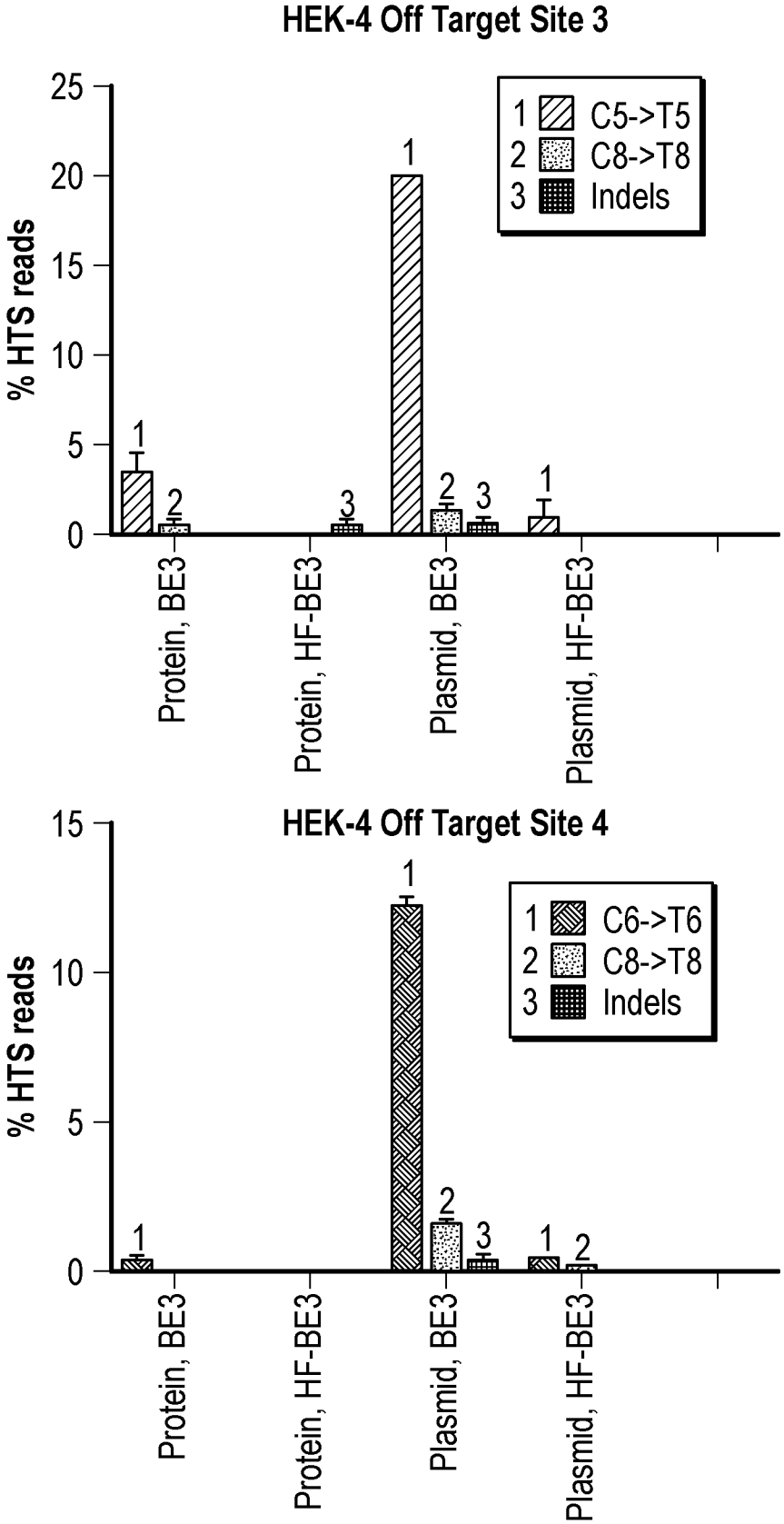


FIGURE 80 (CONTINUED)

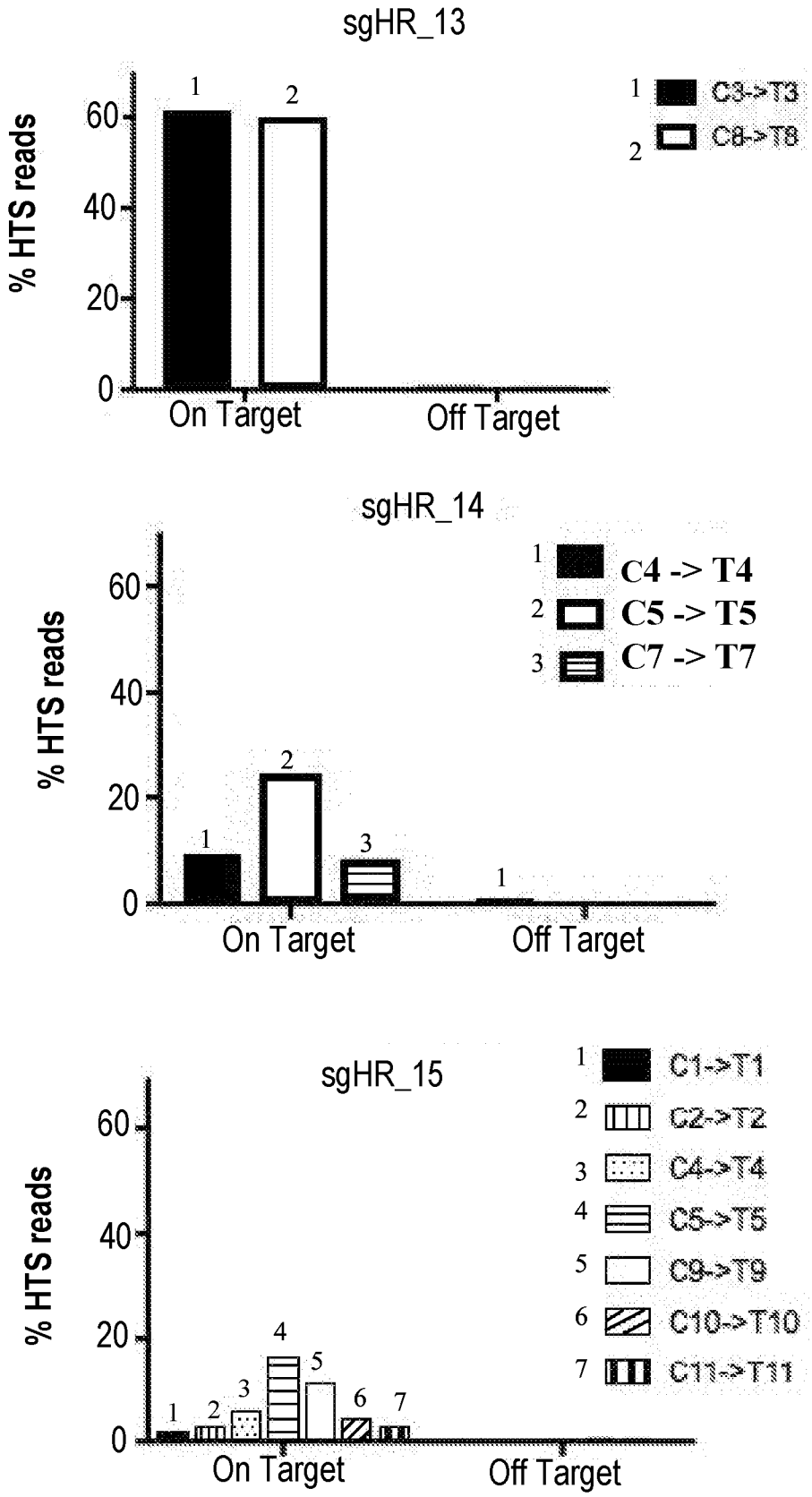

sgHR_13
% HTS reads
60
40
20
0
1
2
On Target
Off Target
1 C3->T3
2 C8->T8
sgHR_14
% HTS reads
60
40
20
0
1
2
3
1
On Target
Off Target
1 C4 -> T4
2 C5 -> T5
3 C7 -> T7
sgHR_15
% HTS reads
60
40
20
0
1
2
3
4
5
6
7
On Target
Off Target
1 C1->T1
2 C2->T2
3 C4->T4
4 C5->T5
5 C9->T9
6 C10->T10
7 C11->T11

FIGURE 81

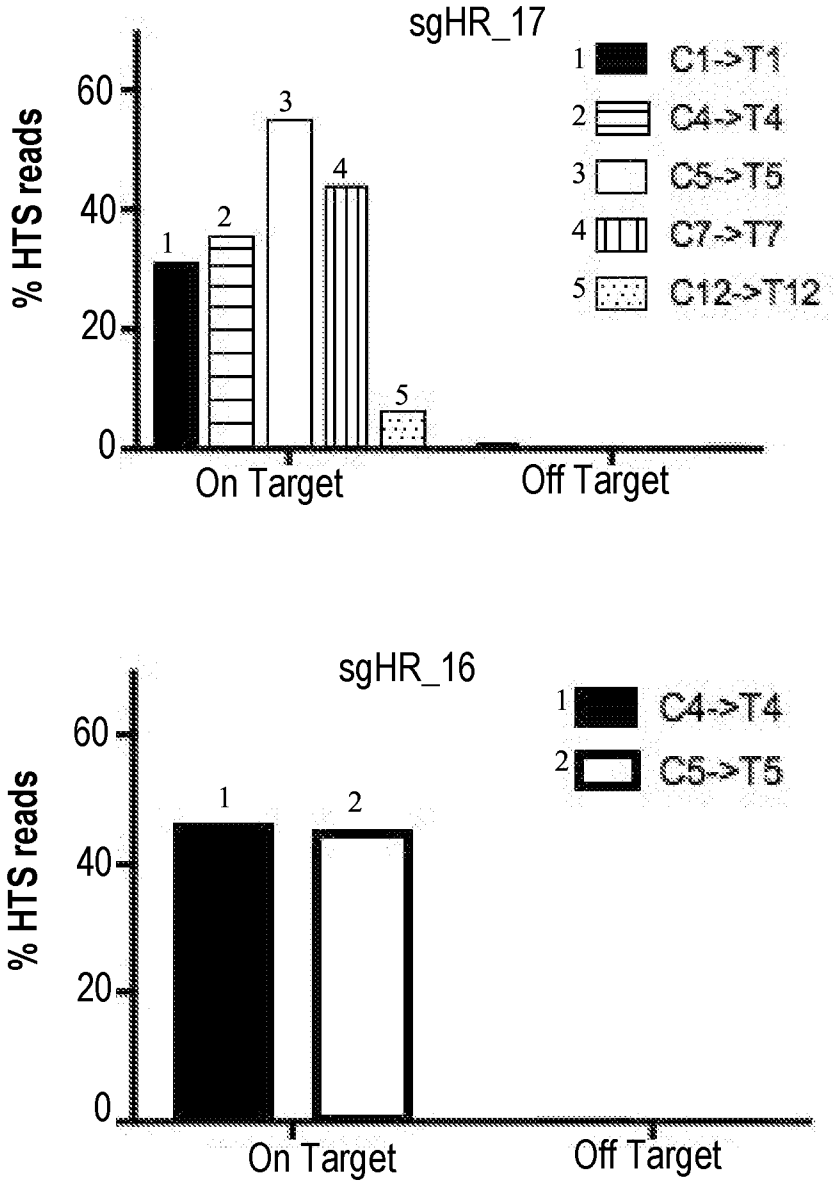

sgHR_17
% HTS reads
60
40
20
0
1
2
3
4
5
On Target
Off Target
1 C1->T1
2 C4->T4
3 C5->T5
4 C7->T7
5 C12->T12
sgHR_16
% HTS reads
60
40
20
0
1
2
On Target
Off Target
1 C4->T4
2 C5->T5

FIGURE 82

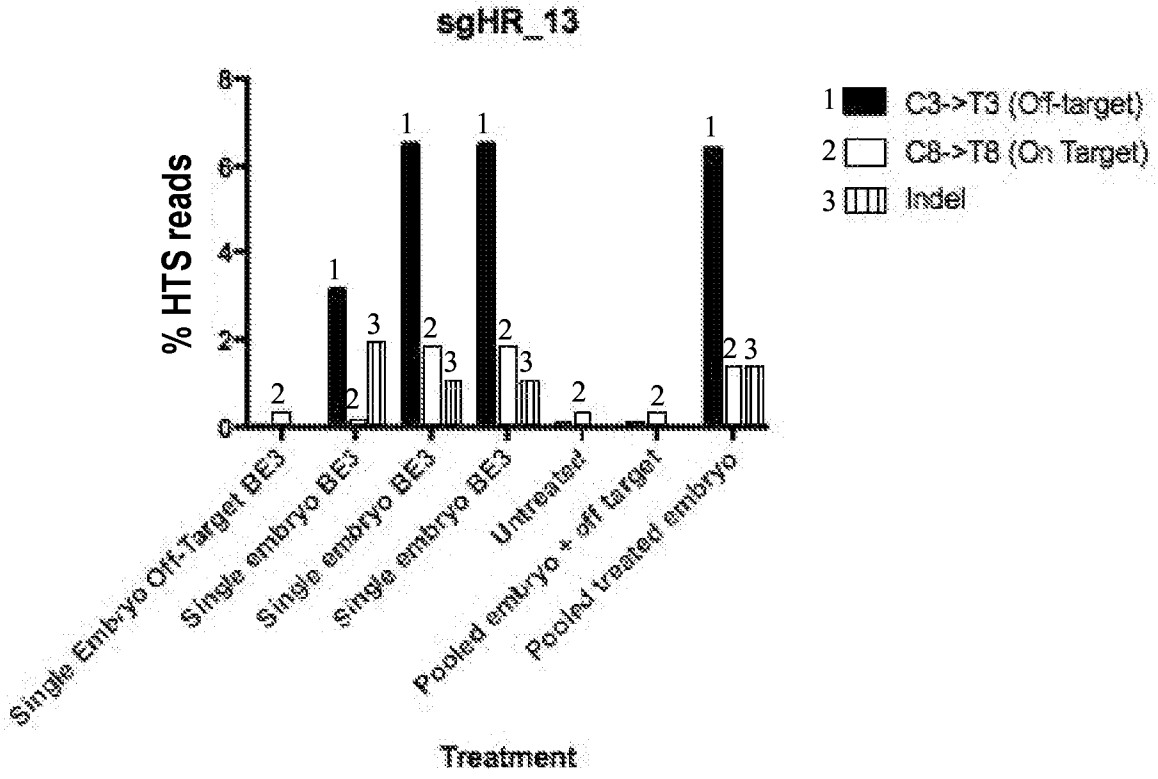


FIGURE 83

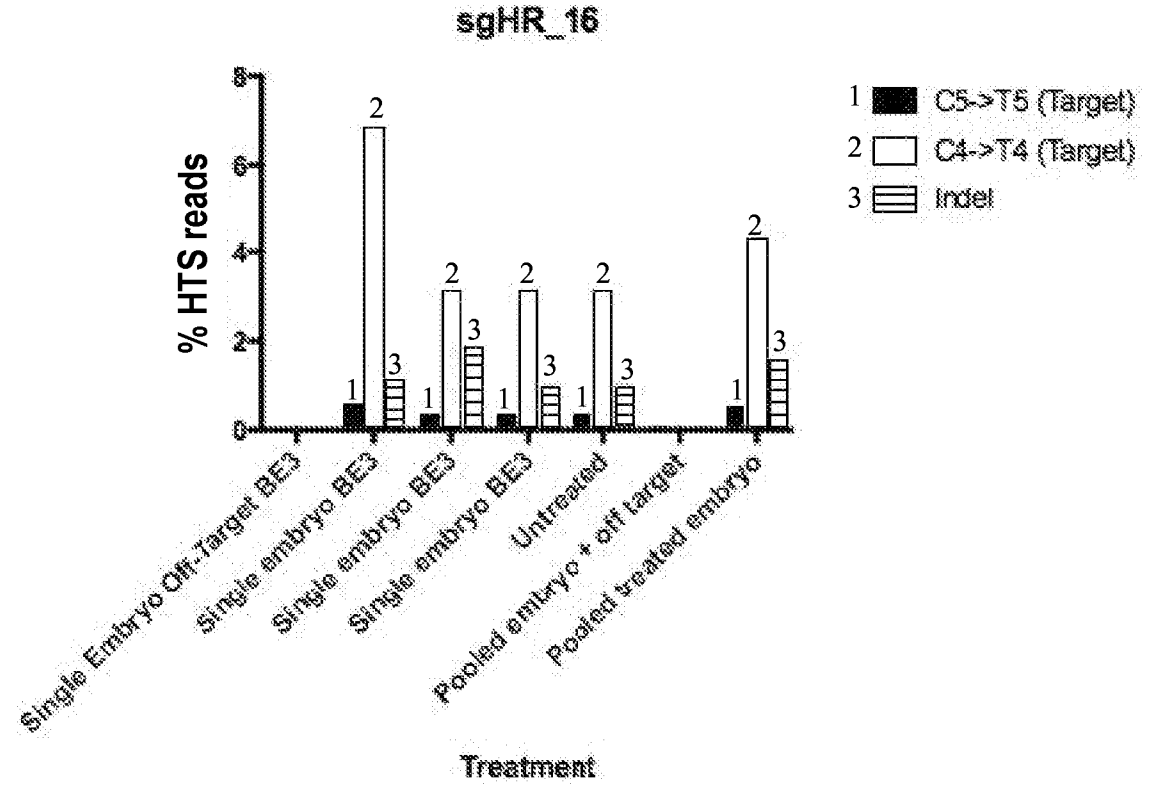


FIGURE 84

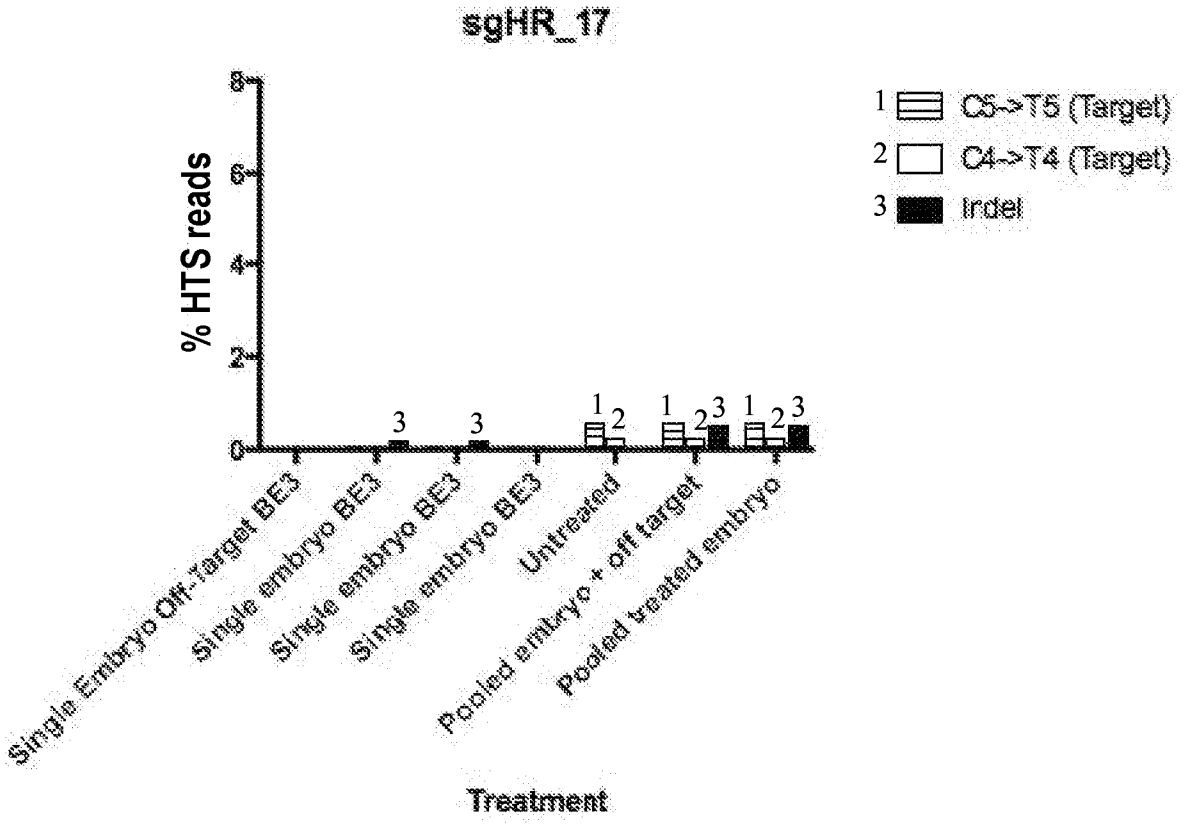

sgHR_17
% HTS reads
0
2
4
6
8
1 C5->T5 (Target)
2 C4->T4 (Target)
3 Indel
3
3
1 2
1 2 3
1 2 3
Single Embryo Off-Target BE3
Single embryo BE3
Single embryo BE3
Single embryo BE3
Untreated
Pooled embryo + off target
Pooled treated embryo
Treatment

FIGURE 85

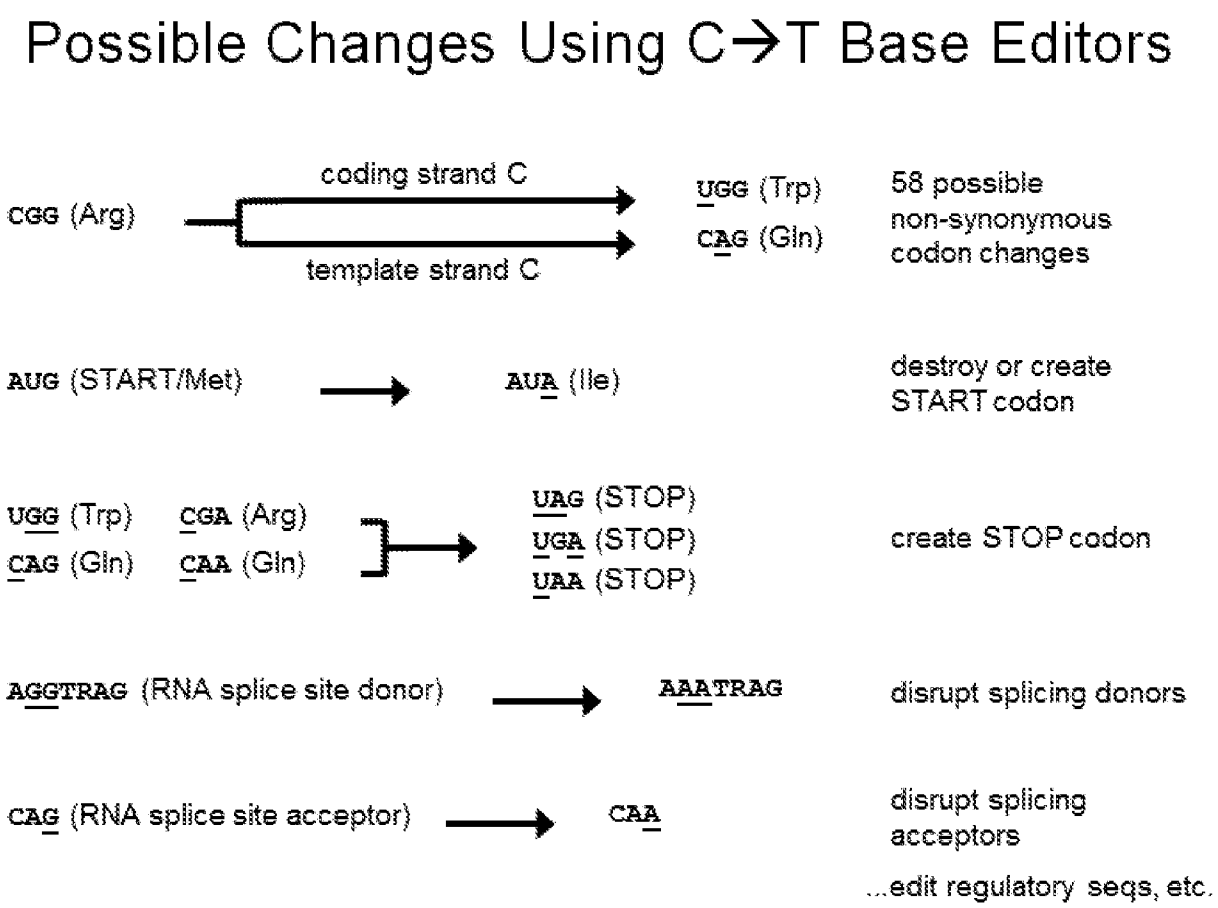
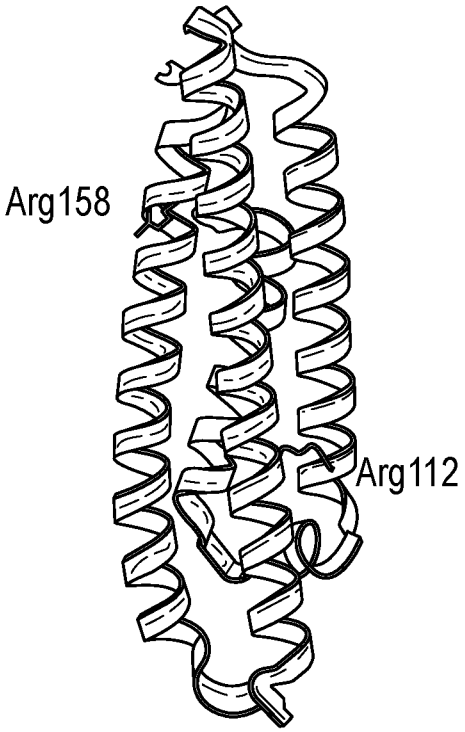

Possible Changes Using C→T Base Editors
CGG (Arg)
coding strand C
template strand C
UGG (Trp)
CAG (Gln)
58 possible non-synonymous codon changes
AUG (START/Met)
AUA (Ile)
destroy or create START codon
UGG (Trp)
CGA (Arg)
CAG (Gln)
CAA (Gln)
UAG (STOP)
UGA (STOP)
UAA (STOP)
create STOP codon
AGGTRAG (RNA splice site donor)
AAATRAG
disrupt splicing donors
CAG (RNA splice site acceptor)
CAA
disrupt splicing acceptors
...edit regulatory seqs, etc.

FIGURE 86

Base Editing *APOE4* to *APOE3r*



- Apolipoprotein E has four isoforms that differ at amino acids 112 and 158
- *APOE4* is the largest and most common genetic risk factor for late-onset Alzheimer's disease

allele	residue 112	residue 158	AD risk (known or presumed)
APOE2	Cys TGC	Cys TGC	low
APOE3	Cys TGC	Arg CGC	neutral
APOE3r	Arg CGC	Cys TGC	neutral
APOE4	Arg CGC	Arg CGC	high

BE3

FIGURE 87

untreated		Lys			Arg 158			Leu			Ala			Val			Tyr			Gln				indel %
APOE4 C158R	G	A	A	G	C_5	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G	**0.0**
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	
C	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

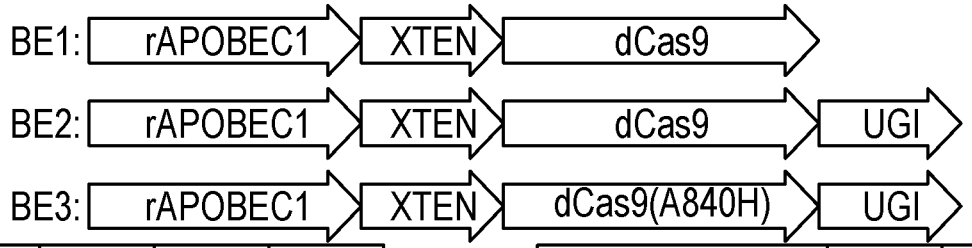
Cas9 + HDR		Lys			Arg 158 → Cys			Leu			Ala			Val			Tyr			Gln				indel%
APOE4 C158R	G	A	A	G	C_5	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G	**26.1**
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	
C	0.0	0.0	0.0	0.0	99.7	0.0	99.9	99.9	0.0	0.0	0.0	100.0	0.5	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	99.6	0.6	99.9	0.2	0.0	0.0	0.0	0.0	100.0	100.0	
T	0.0	0.0	0.0	0.0	**0.3**	0.0	0.1	0.1	100.0	0.0	0.0	0.0	0.1	0.4	99.3	0.1	99.8	0.0	0.0	0.0	0.0	0.0	0.0	

BE3-treated		Lys			Arg 158 → Cys			Leu → Leu			Ala			Val			Tyr			Gln				indel%
APOE4 C158R	G	A	A	G	C_5	G	C_7	C_8	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G	**4.6**
A	0.1	100.0	100.0	0.0	0.5	0.0	1.3	0.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.1	
C	0.0	0.0	0.0	0.0	23.7	0.0	47.4	43.5	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
G	99.9	0.0	0.0	100.0	0.9	99.9	1.1	0.7	0.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	99.9	
T	0.0	0.0	0.0	0.0	**74.9**	0.1	**50.2**	**55.0**	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

FIGURE 88

BE3-treated	Gly				Ser → Ser		Arg 37 → Stop		Tyr			Pro			Gly			Glu			
PRNP R37X	G	G	C	A	G	C_T	C_e	G	A	T	A	C	C	C	G	G	G	C	A	G	G
A	0.0	0.0	0.1	100.0	0.0	1.1	0.3	0.1	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
C	0.0	0.0	98.9	0.0	0.0	54.9	57.8	0.0	0.0	0.0	0.0	99.8	99.8	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
G	100.0	100.0	0.0	0.0	100.0	2.1	0.6	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	100.0
T	0.0	0.0	1.1	0.0	0.0	41.9	41.2	0.0	0.0	100.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

FIGURE 89



EMX1-parental		C_5	C_6	C_{10}	indel %
untreated	A	0.0%	0.0%	0.0%	0.01%
	C	100.0%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	
BE1	A	0.4%	0.1%	0.0%	0.18%
	C	95.8%	97.4%	100.0%	
	G	0.9%	0.2%	0.0%	
	T	2.9%	2.3%	0.0%	
BE2	A	0.3%	0.1%	0.0%	0.15%
	C	94.9%	95.7%	100.0%	
	G	1.0%	0.1%	0.0%	
	T	3.8%	4.0%	0.0%	
BE3	A	1.4%	0.8%	0.0%	2.54%
	C	78.5%	84.7%	100.0%	
	G	3.0%	1.4%	0.0%	
	T	17.1%	13.0%	0.0%	

EMX1-UDG KO		C_5	C_6	C_{10}	indel %
untreated	A	0.0%	0.0%	0.0%	0.01%
	C	99.9%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	0.1%	0.1%	0.0%	
BE1	A	0.0%	0.0%	0.0%	0.04%
	C	76.5%	76.6%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	23.5%	23.4%	0.0%	
BE2	A	0.0%	0.0%	0.0%	0.02%
	C	96.3%	96.3%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	3.7%	3.7%	0.0%	
BE3	A	0.0%	0.0%	0.0%	0.04%
	C	84.5%	86.8%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	15.5%	13.2%	0.0%	

FIGURE 90

FANCF-parental		C_6	C_7	C_8	C_{11}	indel %
untreated	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	99.9%	99.9%	99.9%	99.9%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.1%	0.1%	0.1%	0.0%	
BE1	A	0.5%	0.5%	0.3%	0.0%	0.13%
	C	94.2%	97.8%	97.9%	99.8%	
	G	0.7%	0.0%	0.0%	0.0%	
	T	4.7%	1.6%	1.8%	0.2%	
BE2	A	0.3%	0.4%	0.2%	0.0%	0.25%
	C	95.3%	97.3%	97.6%	99.8%	
	G	0.4%	0.1%	0.0%	0.0%	
	T	4.0%	2.3%	2.1%	0.1%	
BE3	A	2.4%	3.2%	2.2%	0.9%	18.88%
	C	60.3%	72.6%	73.8%	86.6%	
	G	1.2%	0.6%	0.4%	0.3%	
	T	36.2%	23.6%	23.5%	12.3%	

FANCF-UDG KO		C_6	C_7	C_8	C_{11}	indel %
untreated	A	0.0%	0.0%	0.0%	0.0%	0.02%
	C	99.9%	99.9%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.1%	0.0%	0.0%	0.0%	
BE1	A	0.0%	0.1%	0.0%	0.0%	0.09%
	C	60.9%	61.1%	61.3%	94.7%	
	G	0.1%	0.1%	0.1%	0.0%	
	T	39.0%	38.8%	38.6%	5.2%	
BE2	A	0.0%	0.0%	0.0%	0.0%	0.02%
	C	88.6%	88.6%	88.7%	98.9%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	11.4%	11.4%	11.3%	1.1%	
BE3	A	0.0%	0.1%	0.1%	0.0%	0.35%
	C	47.9%	48.5%	48.8%	55.8%	
	G	0.0%	0.1%	0.3%	0.2%	
	T	52.1%	51.3%	50.8%	44.0%	

FIGURE 90 (CONTINUED)

HEK3-parental		C_3	C_4	C_5	C_9	indel %
untreated	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	99.9%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	0.0%	
BE1	A	0.0%	0.4%	0.3%	0.1%	0.07%
	C	99.9%	96.3%	94.4%	99.8%	
	G	0.0%	0.1%	1.9%	0.1%	
	T	0.1%	3.2%	3.4%	0.1%	
BE2	A	0.0%	0.2%	0.3%	0.1%	0.05%
	C	100.0%	97.7%	96.1%	99.8%	
	G	0.0%	0.0%	1.2%	0.1%	
	T	0.0%	2.1%	2.4%	0.1%	
BE3	A	0.1%	2.9%	2.4%	0.3%	3.27%
	C	99.4%	62.6%	55.9%	98.8%	
	G	0.0%	1.5%	10.4%	0.3%	
	T	0.6%	33.0%	31.3%	0.6%	

HEK3-UDG KO		C_3	C_4	C_5	C_9	indel %
untreated	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	100.0%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	0.0%	
BE1	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	96.0%	50.3%	41.1%	96.3%	
	G	0.0%	0.0%	0.1%	0.1%	
	T	4.0%	49.6%	58.8%	3.6%	
BE2	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	99.0%	80.2%	73.2%	98.8%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	1.0%	19.7%	26.8%	1.1%	
BE3	A	0.0%	0.0%	0.0%	0.0%	0.09%
	C	98.5%	55.6%	40.3%	99.4%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	1.4%	44.3%	59.7%	0.6%	

FIGURE 90 (CONTINUED)

HEK4-parental		C_3	C_5	C_8	C_{11}	indel %
untreated	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	100.0%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.1%	0.0%	
BE1	A	0.0%	1.4%	0.0%	0.0%	0.08%
	C	100.0%	95.5%	99.9%	100.0%	
	G	0.0%	2.2%	0.0%	0.0%	
	T	0.0%	0.9%	0.1%	0.0%	
BE2	A	0.0%	0.4%	0.0%	0.0%	0.07%
	C	100.0%	97.7%	99.9%	100.0%	
	G	0.0%	1.1%	0.0%	0.0%	
	T	0.0%	0.7%	0.1%	0.0%	
BE3	A	0.0%	7.1%	0.0%	0.0%	6.05%
	C	99.9%	58.5%	99.7%	100.0%	
	G	0.0%	17.4%	0.0%	0.0%	
	T	0.1%	17.0%	0.2%	0.0%	

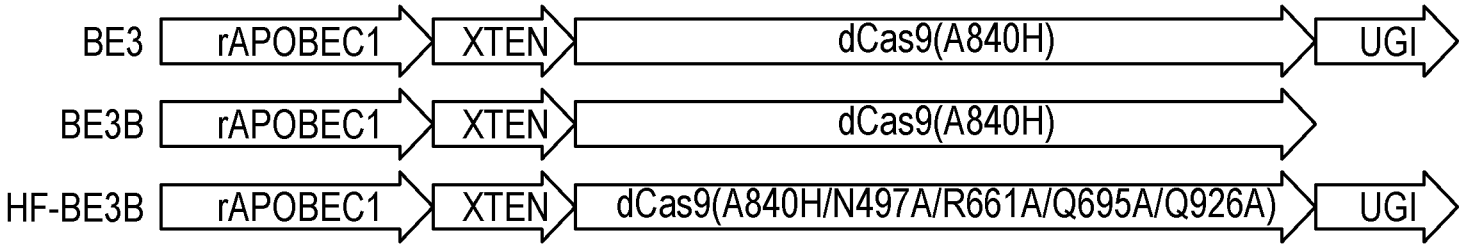
HEK4-UDG KO		C_3	C_5	C_8	C_{11}	indel %
untreated	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	100.0%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.1%	0.0%	
BE1	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	99.9%	67.5%	99.0%	100.0%	
	G	0.0%	0.0%	0.1%	0.0%	
	T	0.1%	32.5%	1.0%	0.0%	
BE2	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	91.9%	99.7%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	8.1%	0.2%	0.0%	
BE3	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	99.9%	65.9%	99.7%	100.0%	
	G	0.0%	0.3%	0.0%	0.0%	
	T	0.1%	33.8%	0.3%	0.0%	

FIGURE 90 (CONTINUED)

EMX1		C_3	C_6	C_{12}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	
BE3	A	1.3 ± 0.1	0.7 ± 0.2	0 ± 0	2.6 ± 0.3
	C	51.7 ± 1	55.2 ± 0.1	99.9 ± 0	
	G	4 ± 0.7	1.9 ± 0.4	0 ± 0	
	T	42.8 ± 0.2	41.9 ± 0.4	0 ± 0	
BE3B	A	4.9 ± 1.4	3.9 ± 1.1	0 ± 0	11.6 ± 2.6
	C	52.6 ± 12.1	61.9 ± 10.5	99.9 ± 0	
	G	12.3 ± 3.3	11.3 ± 3.1	0 ± 0	
	T	30.1 ± 7.3	22.7 ± 6.1	0 ± 0	

FANCF		C_6	C_7	C_8	C_{11}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0.8 ± 0.1	1.2 ± 0.3	1.2 ± 0.1	0.2 ± 0.1	5.8 ± 0.9
	C	64.7 ± 5.9	67.9 ± 5.1	68.9 ± 5.2	85.4 ± 2.7	
	G	0.6 ± 0	0.7 ± 0	0.7 ± 0.1	0.1 ± 0	
	T	33.8 ± 5.7	29.9 ± 4.8	29 ± 5	14.1 ± 2.6	
BE3B	A	2.9 ± 0.9	4.1 ± 1.4	2.8 ± 0.9	1.1 ± 0	20.5 ± 3.8
	C	67.3 ± 8.8	79.3 ± 5.8	79.5 ± 5.2	92 ± 1.9	
	G	2 ± 0.5	0.7 ± 0.3	0.9 ± 0.2	0.5 ± 0.1	
	T	27.7 ± 7.3	15.6 ± 4	16.5 ± 4	6.1 ± 1.3	

FIGURE 91



HEK3		C_6	C_7	C_8	C_{11}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0 ± 0	1.6 ± 0.2	1.5 ± 0.3	0.2 ± 0	2.7 ± 0.4
	C	98.7 ± 0.2	48.6 ± 5.4	40.8 ± 6.7	98.6 ± 0.1	
	G	0 ± 0	1.4 ± 0.2	10.5 ± 1.2	0.3 ± 0	
	T	1.1 ± 0.2	48.2 ± 4.9	47 ± 5.2	0.7 ± 0.1	
BE3B	A	0.1 ± 0	4.9 ± 0.8	3.8 ± 0.8	0.8 ± 0.1	7.7 ± 1
	C	99.4 ± 0.1	57.6 ± 6.3	44.4 ± 8.4	97.7 ± 0.3	
	G	0 ± 0	3.3 ± 0.7	21.2 ± 2.8	0.7 ± 0.1	
	T	0.4 ± 0.1	33.9 ± 4.7	30.4 ± 4.7	0.6 ± 0	

HEK4		C_6	C_7	C_8	C_{11}	indel %
untreated	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0 ± 0	5.4 ± 1.3	0 ± 0	0 ± 0	3 ± 0.7
	C	98.7 ± 0	44.5 ± 8.2	98.9 ± 0	99.8 ± 0	
	G	0 ± 0	18.1 ± 2.3	0 ± 0	0 ± 0	
	T	1.1 ± 0.1	31.9 ± 4.8	0.9 ± 0.1	0.1 ± 0	
BE3B	A	0 ± 0	10.6 ± 2.6	0.3 ± 0	0 ± 0	7.6 ± 0.7
	C	99.6 ± 0	45 ± 9.9	99.2 ± 0.1	99.9 ± 0	
	G	0 ± 0	32.8 ± 5.8	0 ± 0	0 ± 0	
	T	0.2 ± 0	11.4 ± 1.6	0.4 ± 0.2	0 ± 0	

FIGURE 91 (CONTINUED)

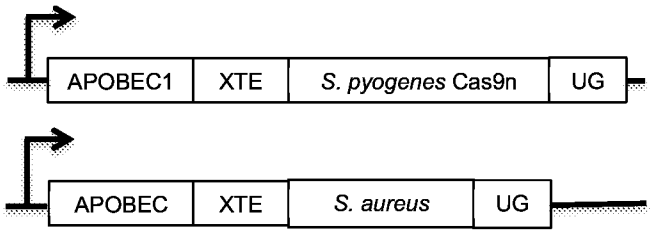


FIGURE 92A

Species	PAM	Base editor	Reference
S. pyogenes	NGG	BE3	Wild-type
	NGA	VQR, EQR BE3	Ref #7
	NGCG	VRER BE3	Ref #7
S. aureus	NNGRRT	SaBE3	Wild-type
	NNNRRT	SaKKHBE3	Ref #8

FIGURE 92B

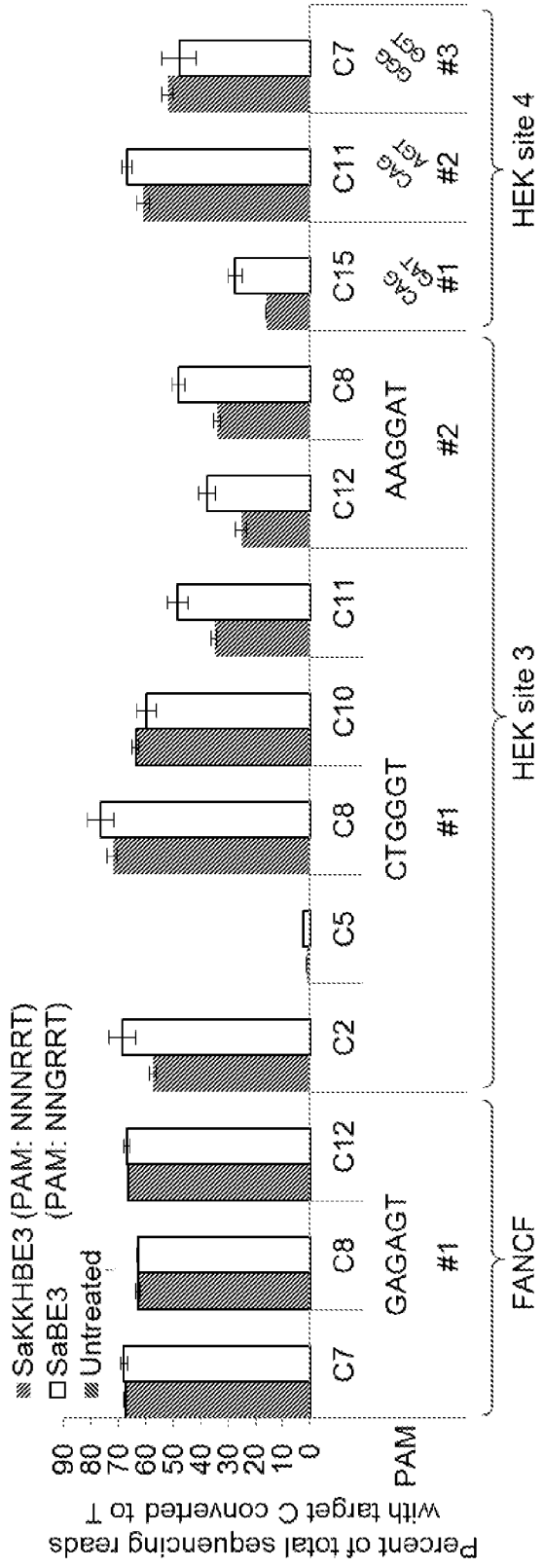

SaKKHBE3 (PAM: NNNRRT)
SaBE3 (PAM: NNGRRT)
Untreated
Percent of total sequencing reads with target C converted to T
90
80
70
60
50
40
30
20
10
0
PAM
C7
C8
C12
C2
C5
C8
C10
C11
C12
C8
C15
C11
C7
GAGAGT
CTGGGT
AAGGAT
CAG GAT
CAG ACT
GGG GGT
#1
#1
#2
#1
#2
#3
FANCF
HEK site 3
HEK site 4

FIGURE 92C

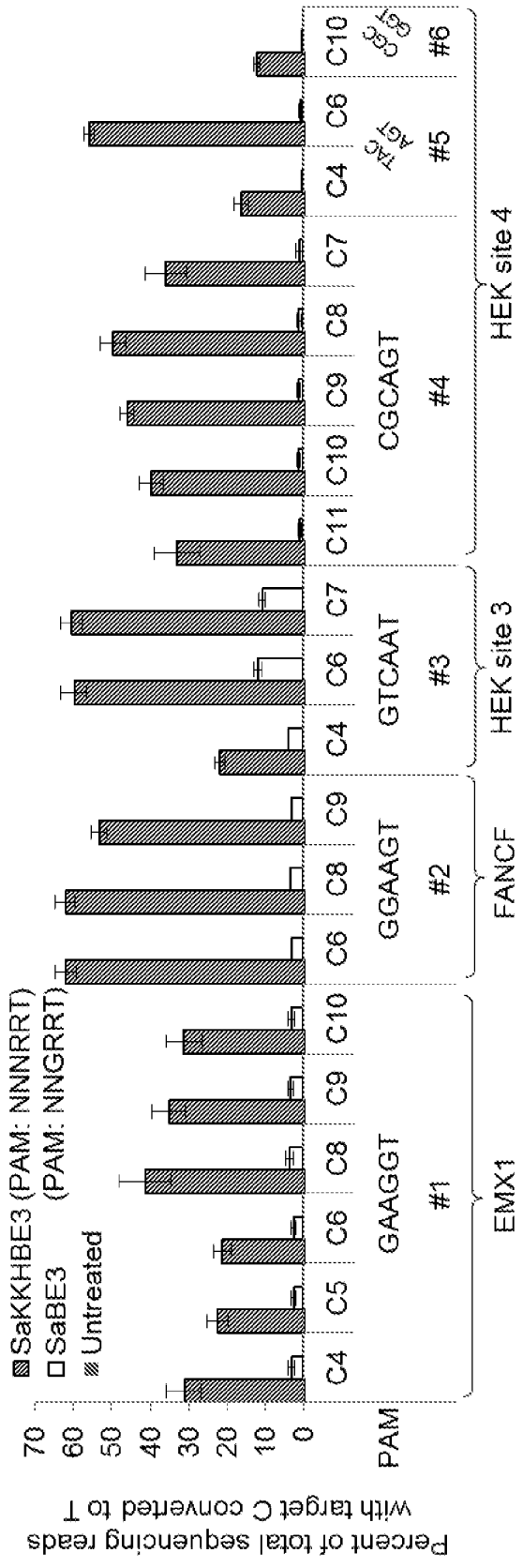

SaKKHBE3 (PAM: NNNRRT)
SaBE3 (PAM: NNGRRT)
Untreated
Percent of total sequencing reads with target C converted to T
70
60
50
40
30
20
10
0
PAM
C4
C5
C6
C8
C9
C10
GAAGGT
#1
EMX1
C6
C8
C9
GGAAGT
#2
FANCF
C4
C6
C7
GTCAAT
#3
HEK site 3
C7
C8
C9
C10
C11
CGCAGT
#4
HEK site 4
C4
C6
AGTAGC
#5
C10
GGCGGT
#6

FIGURE 92D

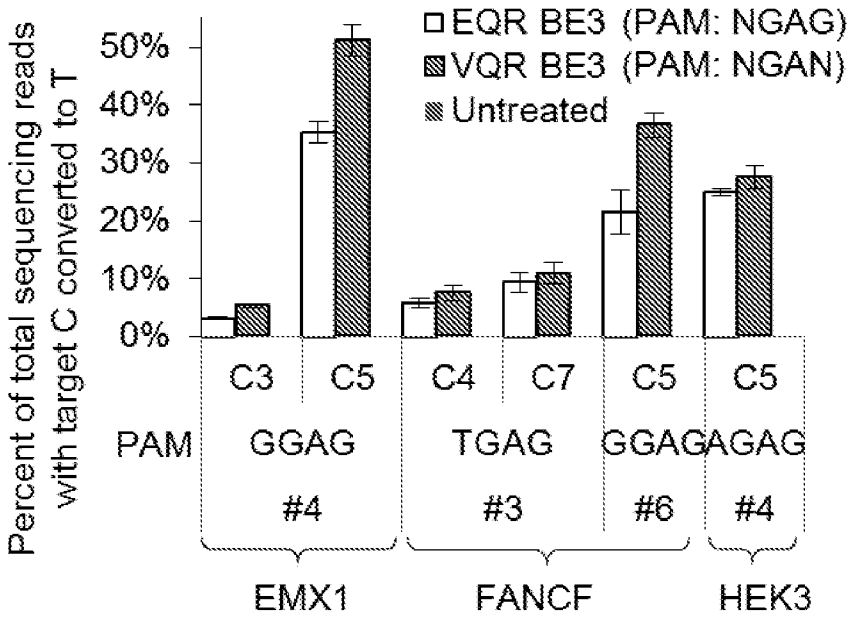


FIGURE 92E

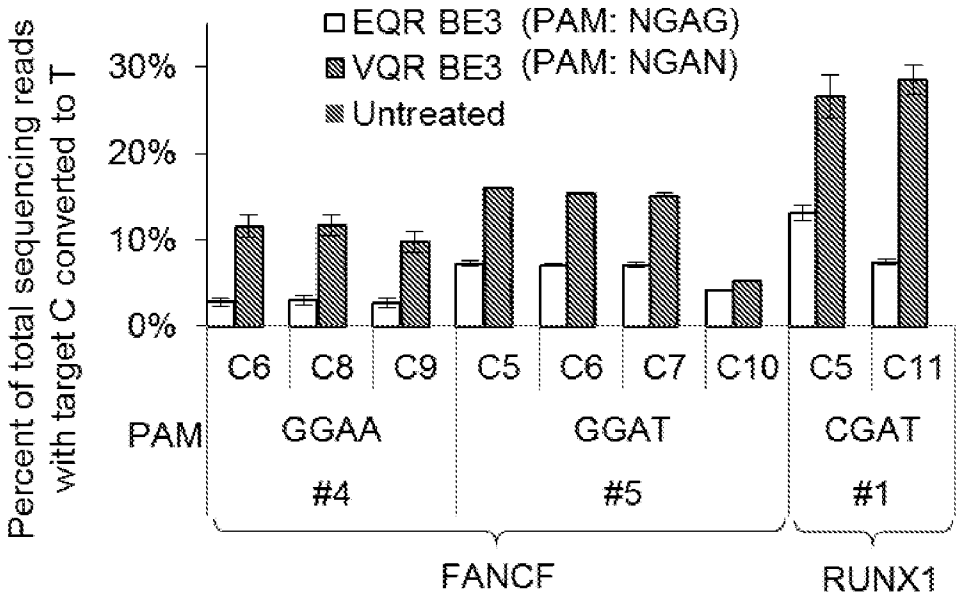


FIGURE 92F

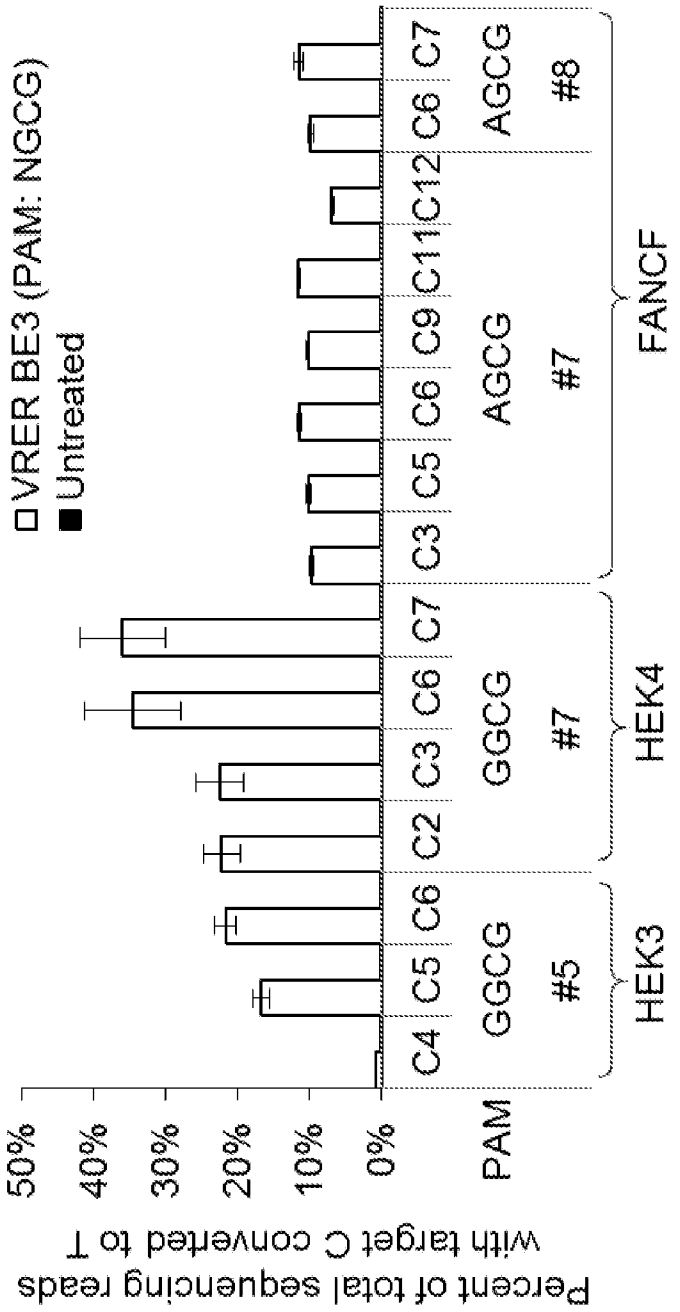

VRER BE3 (PAM: NGCG)
Untreated
Percent of total sequencing reads with target C converted to T
50%
40%
30%
20%
10%
0%
PAM
C4 C5 C6 C2 C3 C6 C7 C3 C5 C6 C9 C11 C12 C6 C7
GGCG GGCG AGCG AGCG
#5 #7 #7 #8
HEK3 HEK4 FANCF

FIGURE 92G

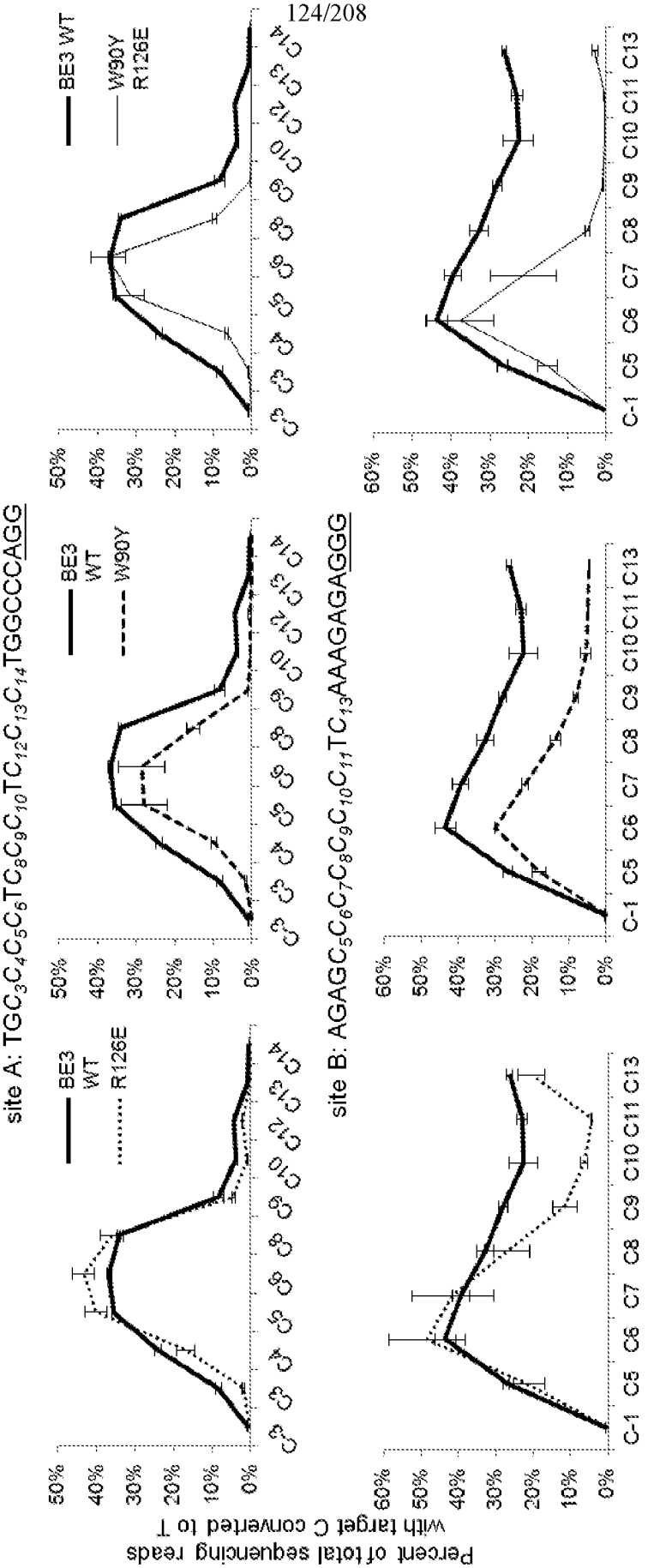

site A: TGC3C4C5C6TC8C9C10TC12C13C14TGGCCCAGG
site B: AGAGC5C6C7C8C9C10C11TC13AAAGAGAGGG
BE3
WT
R126E
BE3
WT
W90Y
BE3 WT
W90Y
R126E
Percent of total sequencing reads with target C converted to T
C-1 C5 C6 C7 C8 C9 C10 C11 C13
0% 10% 20% 30% 40% 50% 60%

FIGURE 93A

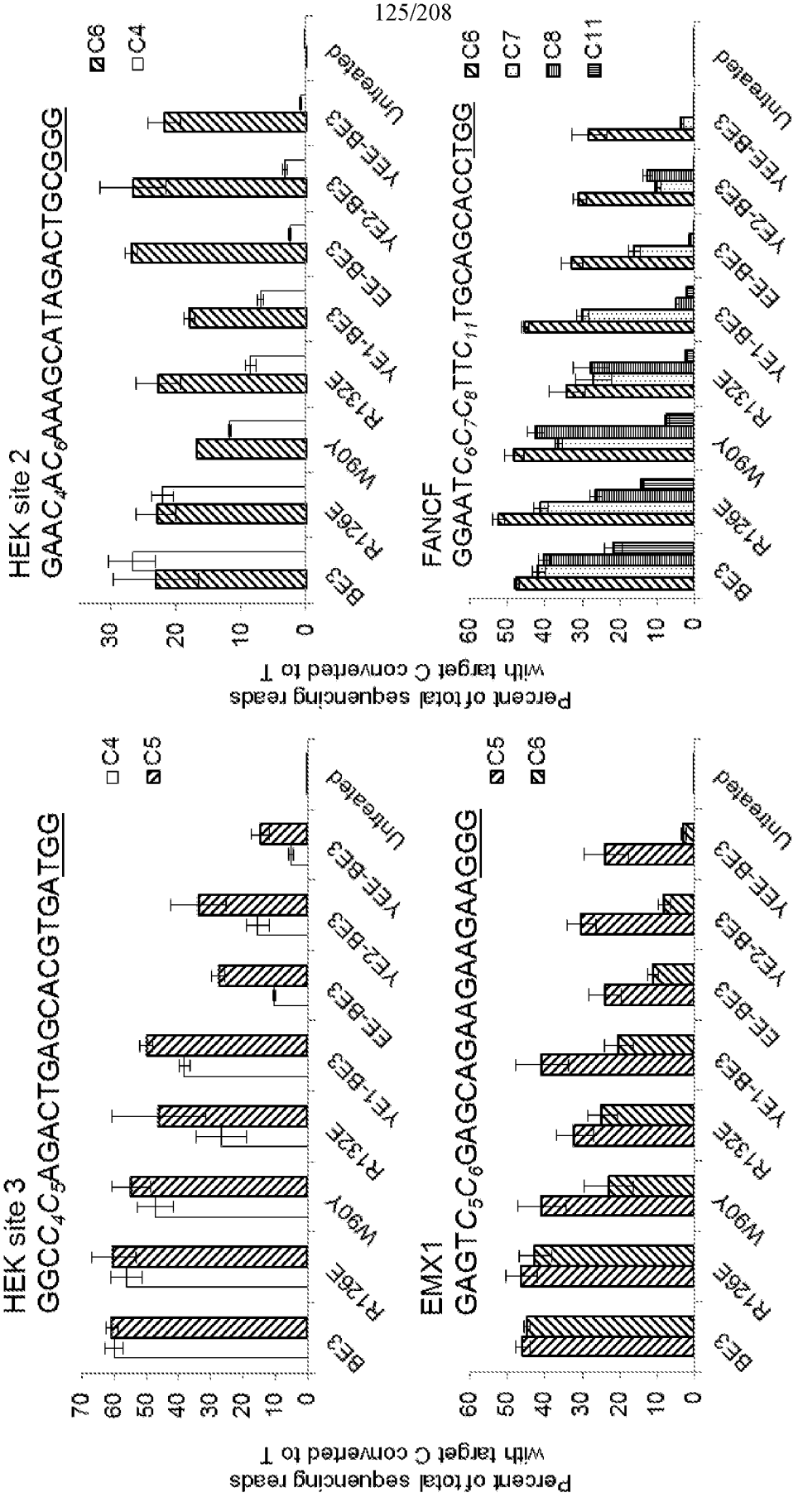

HEK site 3
GGCC4C5AGACTGAGCACGTGATGG
EMX1
GAGTC5C6GAGCAGAAGAAGAAGGG
HEK site 2
GAAC4AC6AAAGCATAGACTGCGGG
FANCF
GGAATC6C7C8TTC11TGCAGCACCTGG
Percent of total sequencing reads with target C converted to T
BE3
R126E
W90Y
R132E
YE1-BE3
EE-BE3
YE2-BE3
YEE-BE3
Untreated
C4
C5
C6
C7
C8
C11

FIGURE 93B

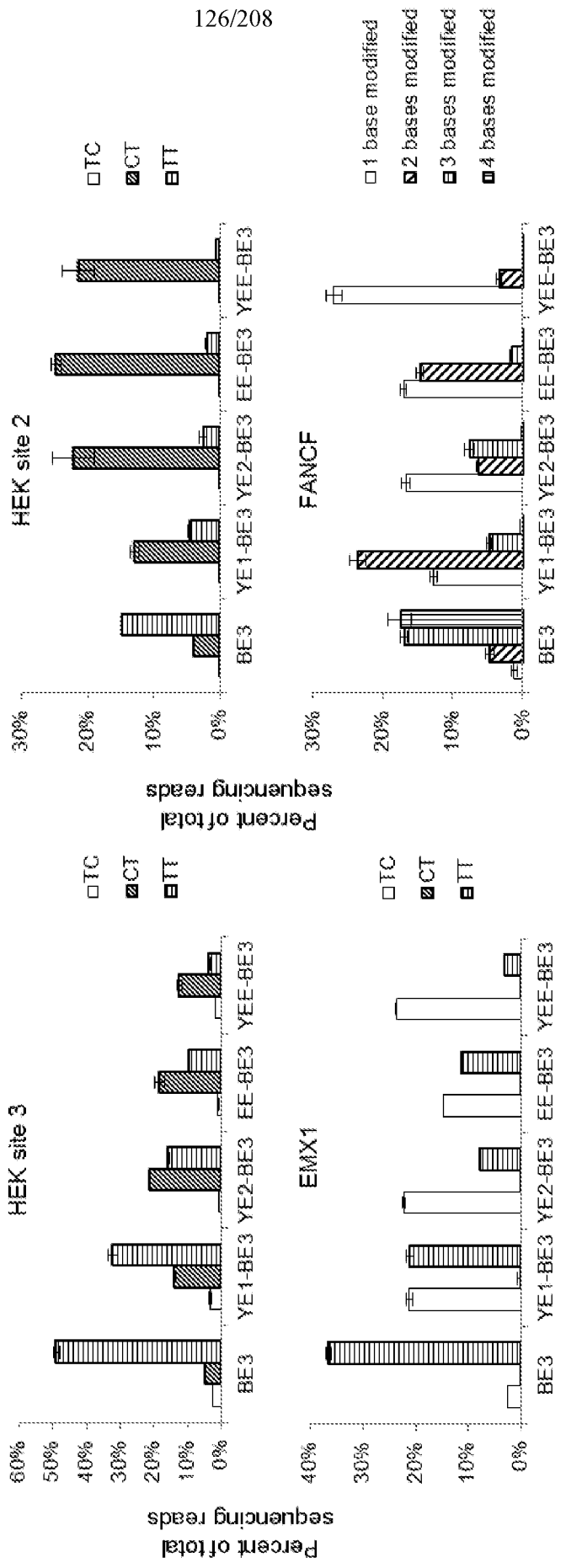

HEK site 3
EMX1
HEK site 2
FANCF
Percent of total sequencing reads
TC
CT
TT
1 base modified
2 bases modified
3 bases modified
4 bases modified
BE3
YE1-BE3
YE2-BE3
EE-BE3
YEE-BE3

FIGURE 93C

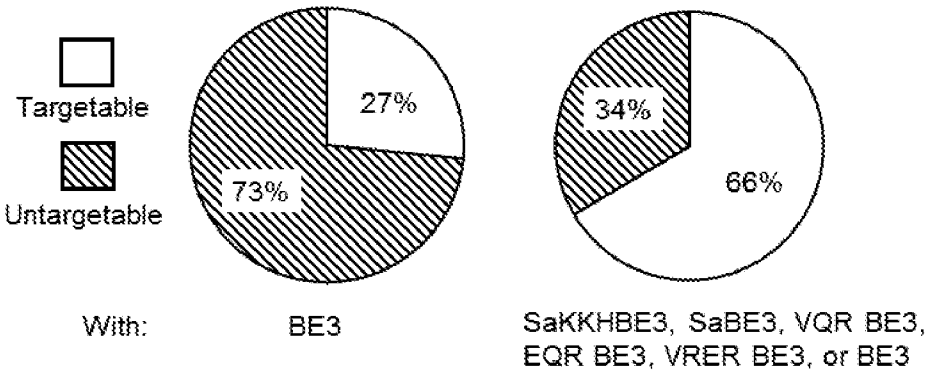


FIGURE 94A

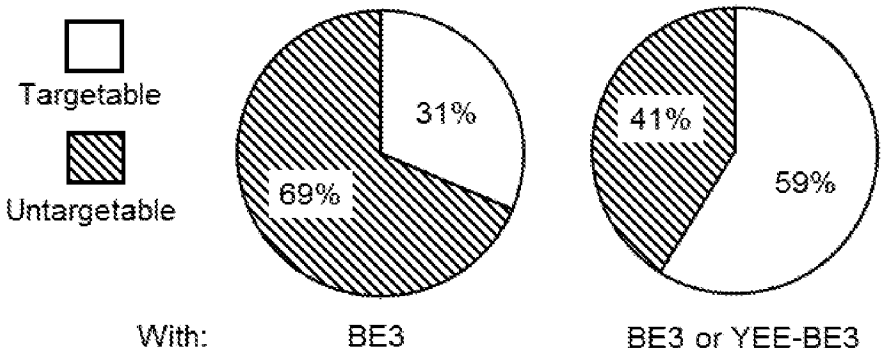


FIGURE 94B

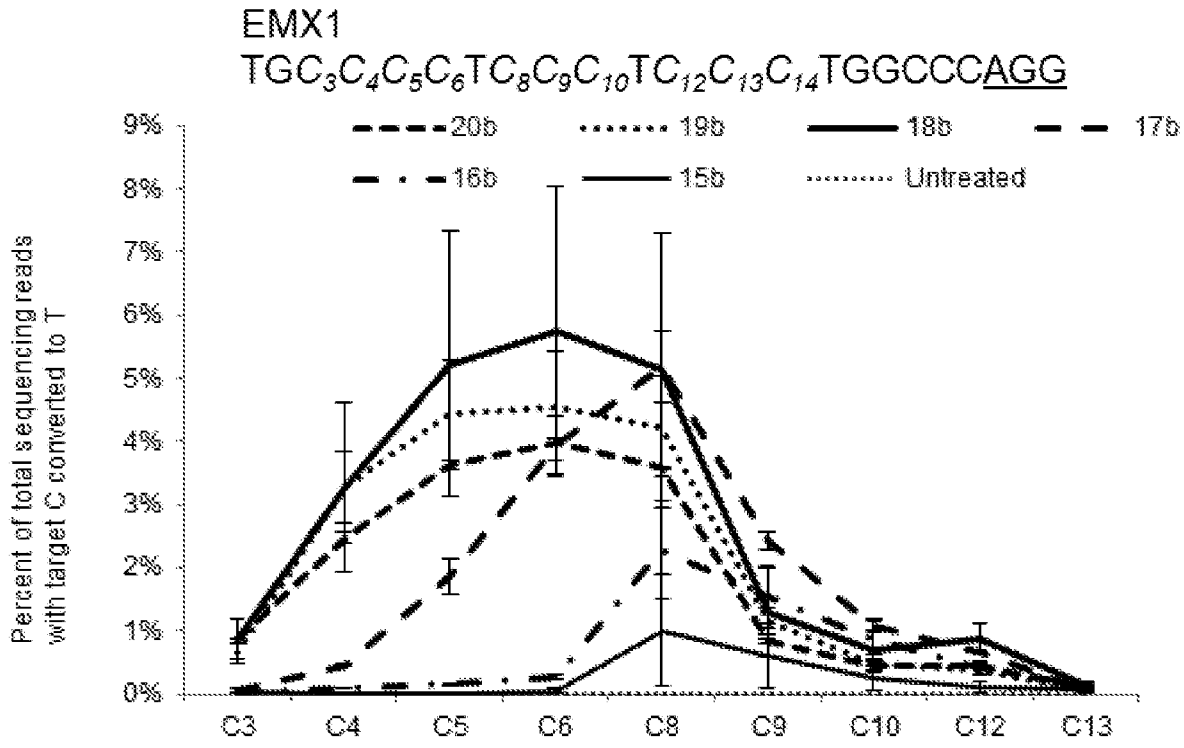

EMX1
TGC3C4C5C6TC8C9C10TC12C13C14TGGCCCAGG
20b
19b
18b
17b
16b
15b
Untreated
Percent of total sequencing reads with target C converted to T
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
C3
C4
C5
C6
C8
C9
C10
C12
C13

FIGURE 95A

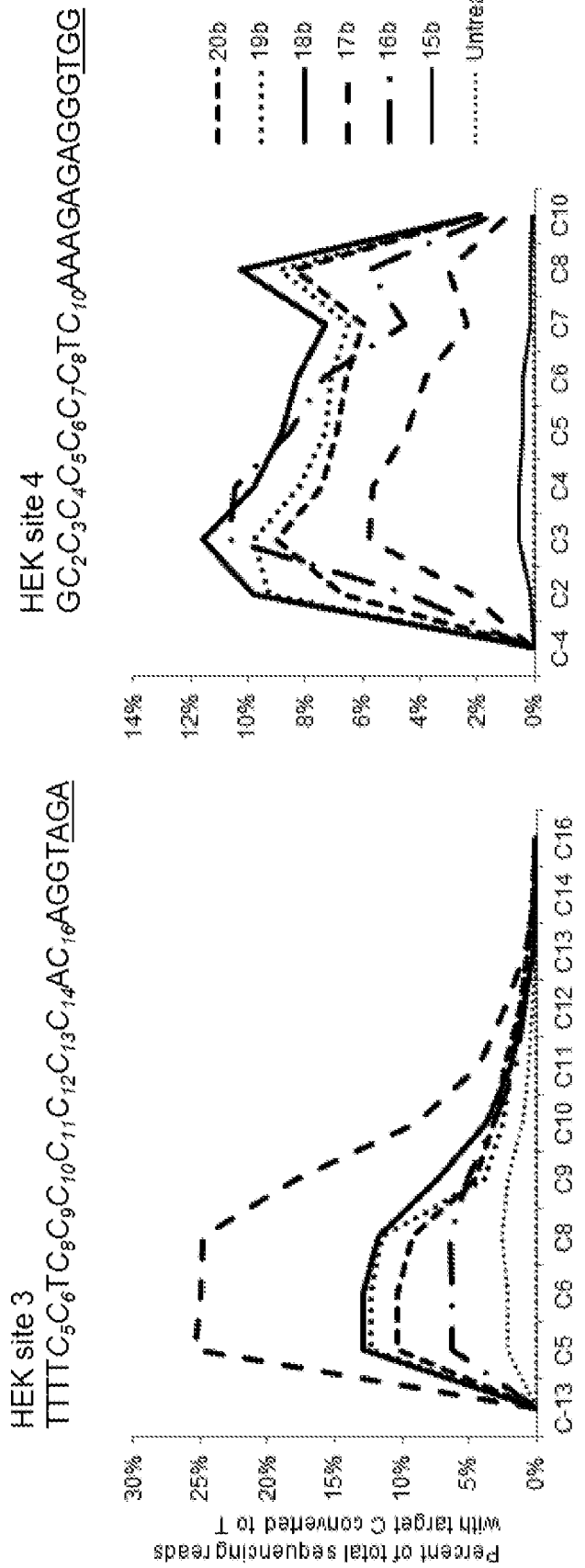

HEK site 3
TTTTC5C6TC8C9C10C11C12C13C14AC16AGGTAGA
HEK site 4
GC2C3C4C5C6C7C8TC10AAAGAGAGGGTGG
Percent of total sequencing reads with target C converted to T
30%
25%
20%
15%
10%
5%
0%
14%
12%
10%
8%
6%
4%
2%
0%
C-13 C5 C6 C8 C9 C10 C11 C12 C13 C14 C16
C-4 C2 C3 C4 C5 C6 C7 C8 C10
20b
19b
18b
17b
16b
15b
Untreated

FIGURE 95B

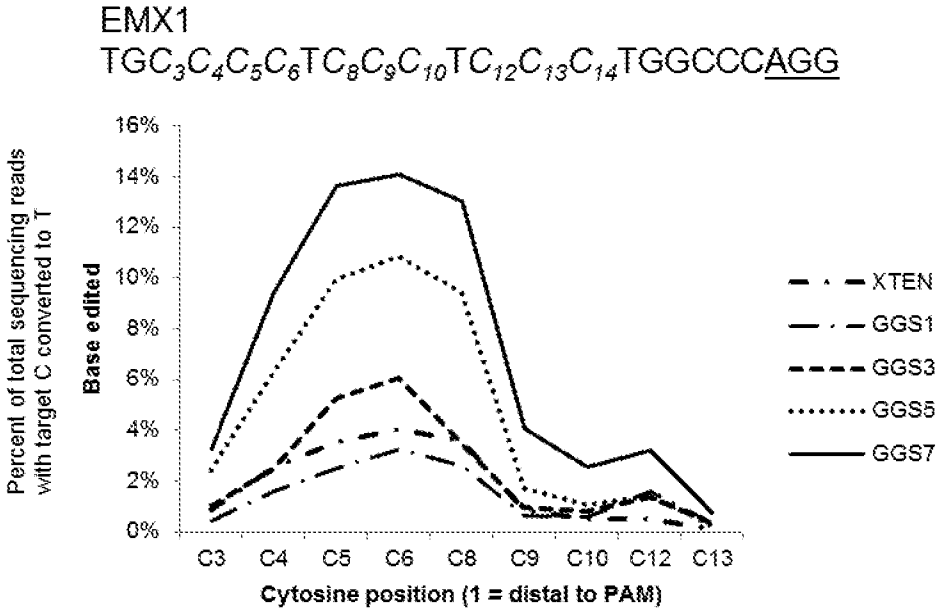


FIGURE 96

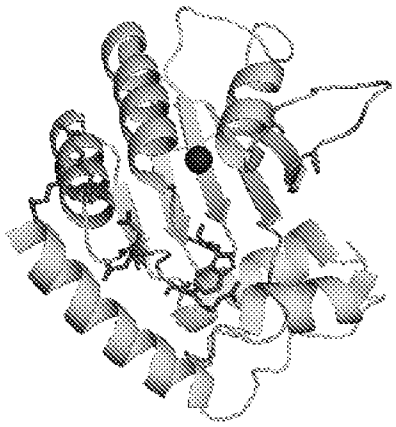


FIGURE 97A

APOBEC1 mutation	APOBEC3G mutation	Reference
R126A	R320A	#9,10
R126E	R320E	#9,10
W90A	W285A	#9,10
W90Y	W285Y	This work
R132E	R326E	This work

FIGURE 97B

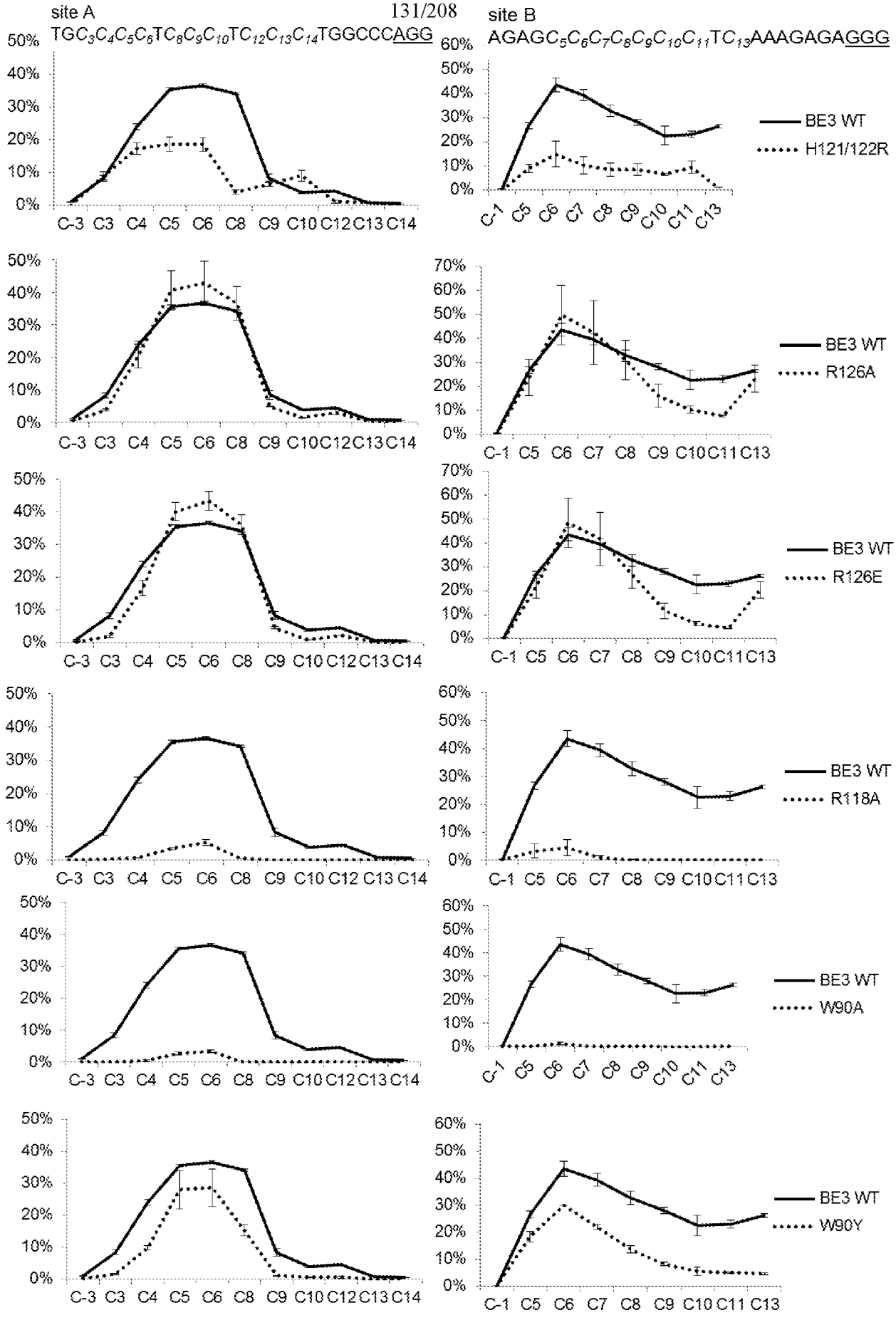
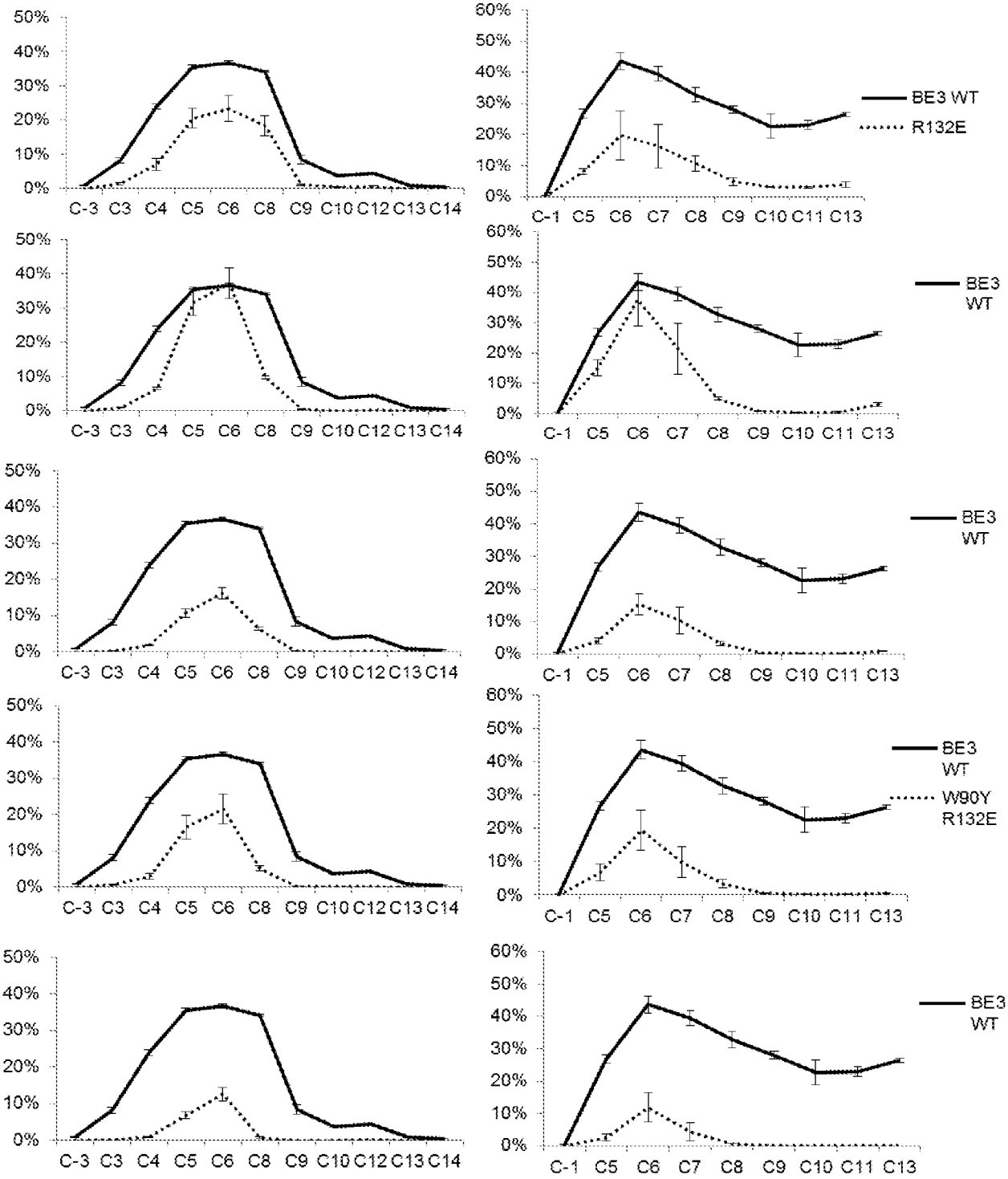

site A
TGC3C4C5C6TC8C9C10TC12C13C14TGGCCCAGG
site B
AGAGC5C6C7C8C9C10C11TC13AAAGAGAGGG
BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
C-3 C3 C4 C5 C6 C8 C9 C10 C12 C13 C14
C-1 C5 C6 C7 C8 C9 C10 C11 C13

FIGURE 97C



Editing window width

	Site A	Site B		Site A	Site B
BE3 WT	5.0	6.1	W90Y	3.8	4.9
H121/122R	4.2	7.4	R132E	4.0	3.0
R126A	4.4	3.4	W90Y R126E	2.9	3.0
R126E	4.2	3.1	R126E R132E	2.9	3.0
R118A	2.4	3.6	W90Y R132E	2.7	2.8
W90A	2.5	1.1	W90Y R126E R132E	2.1	1.4

FIGURE 97C (CONTINUED)

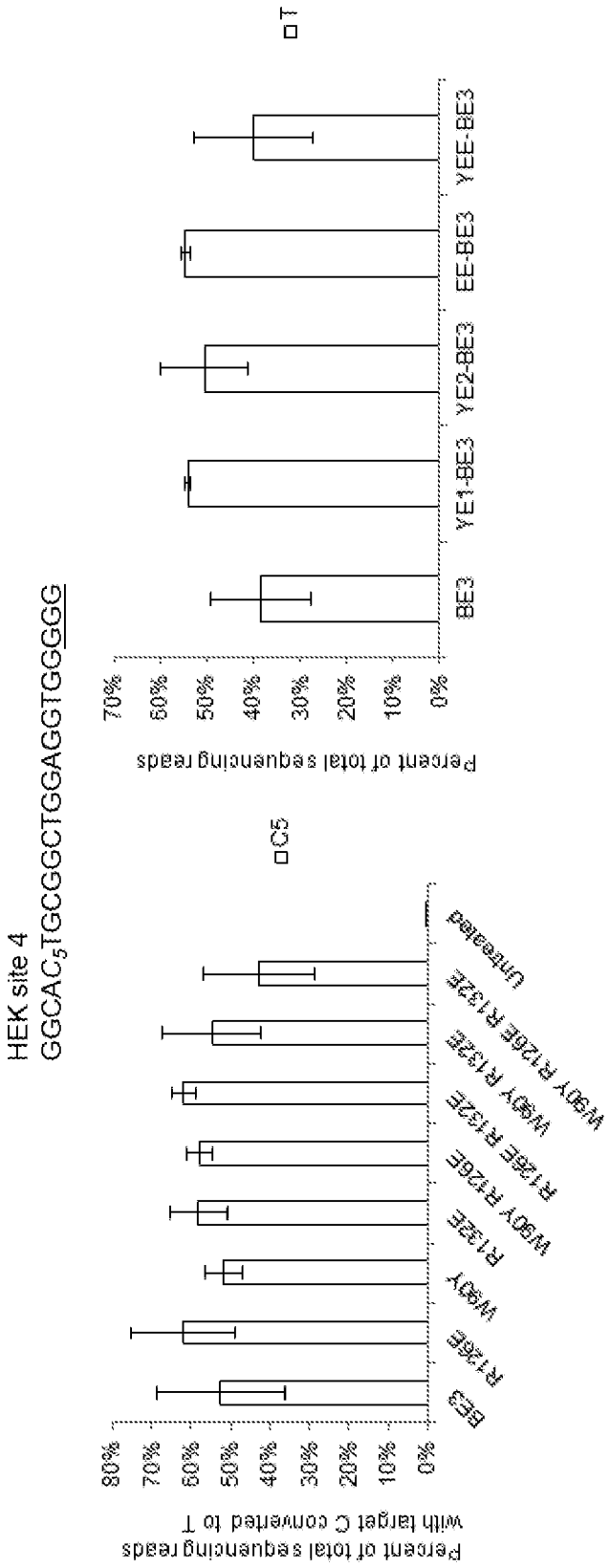

HEK site 4
GGCAC5TGCGGCTGGAGGTGGGGG
C
T
Percent of total sequencing reads with target C converted to T
Percent of total sequencing reads
BE3
R126E
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
unTreated
BE3
YE1-BE3
YE2-BE3
EE-BE3
YEE-BE3

FIGURE 98

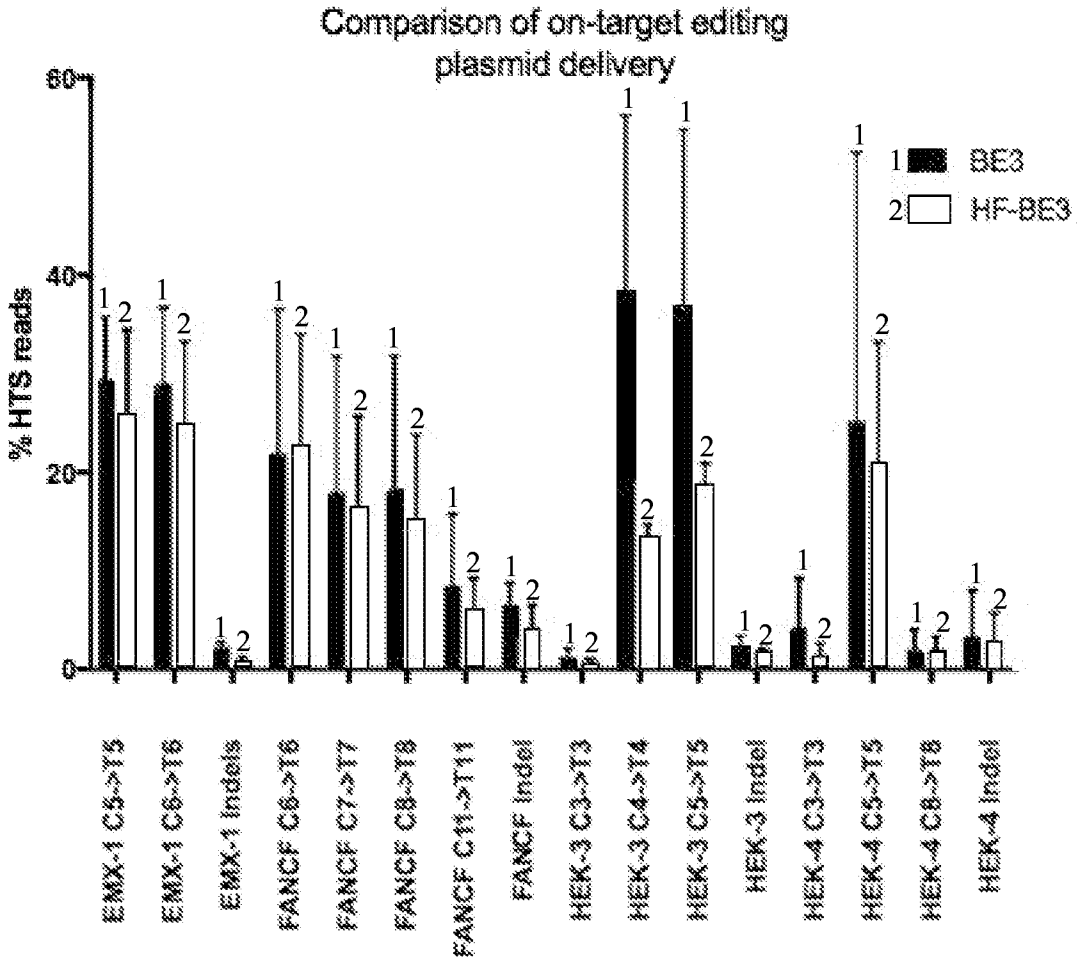

Comparison of on-target editing
plasmid delivery
% HTS reads
60
40
20
0
1 BE3
2 HF-BE3
EMX-1 C5->T5
EMX-1 C6->T6
EMX-1 Indels
FANCF C6->T6
FANCF C7->T7
FANCF C8->T8
FANCF C11->T11
FANCF Indel
HEK-3 C3->T3
HEK-3 C4->T4
HEK-3 C5->T5
HEK-3 Indel
HEK-4 C3->T3
HEK-4 C5->T5
HEK-4 C8->T8
HEK-4 Indel

FIGURE 99

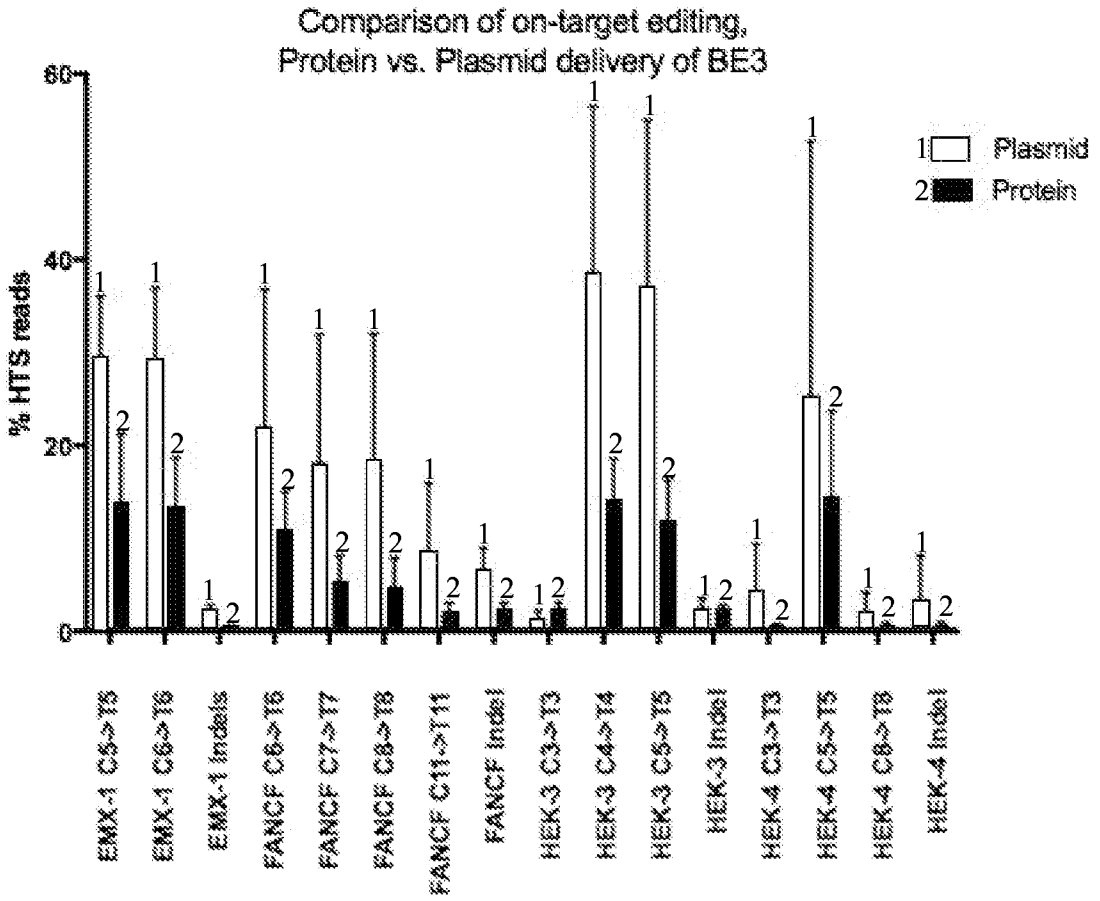

Comparison of on-target editing,
Protein vs. Plasmid delivery of BE3
% HTS reads
60
40
20
0
1 Plasmid
2 Protein
EMX-1 C5->T5
EMX-1 C6->T6
EMX-1 Indels
FANCF C6->T6
FANCF C7->T7
FANCF C8->T8
FANCF C11->T11
FANCF Indel
HEK-3 C3->T3
HEK-3 C4->T4
HEK-3 C5->T5
HEK-3 Indel
HEK-4 C3->T3
HEK-4 C5->T5
HEK-4 C8->T8
HEK-4 Indel

FIGURE 100

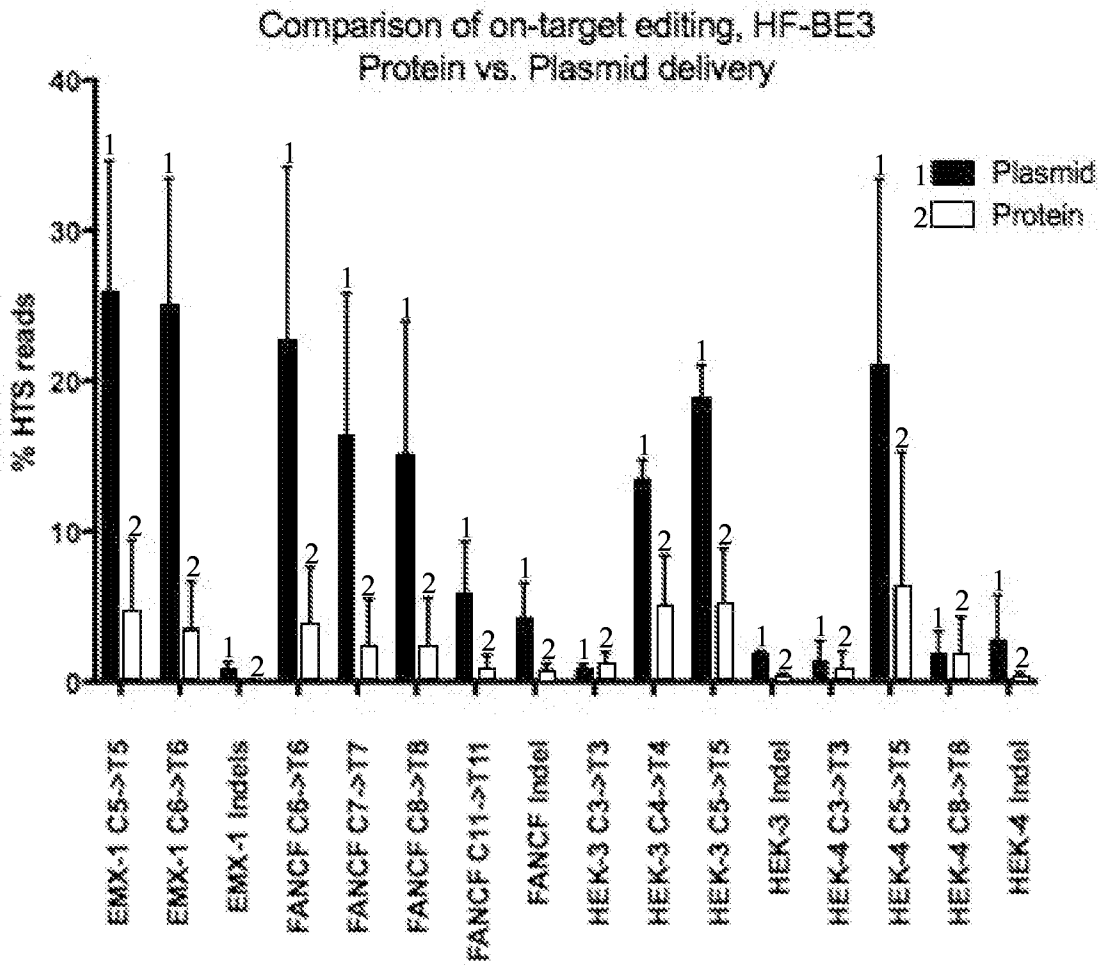

Comparison of on-target editing, HF-BE3
Protein vs. Plasmid delivery
% HTS reads
0
10
20
30
40
1 Plasmid
2 Protein
EMX-1 C5->T5
EMX-1 C6->T8
EMX-1 Indels
FANCF C6->T6
FANCF C7->T7
FANCF C8->T8
FANCF C11->T11
FANCF Indel
HEK-3 C3->T3
HEK-3 C4->T4
HEK-3 C5->T5
HEK-3 Indel
HEK-4 C3->T3
HEK-4 C5->T5
HEK-4 C8->T8
HEK-4 Indel

FIGURE 101

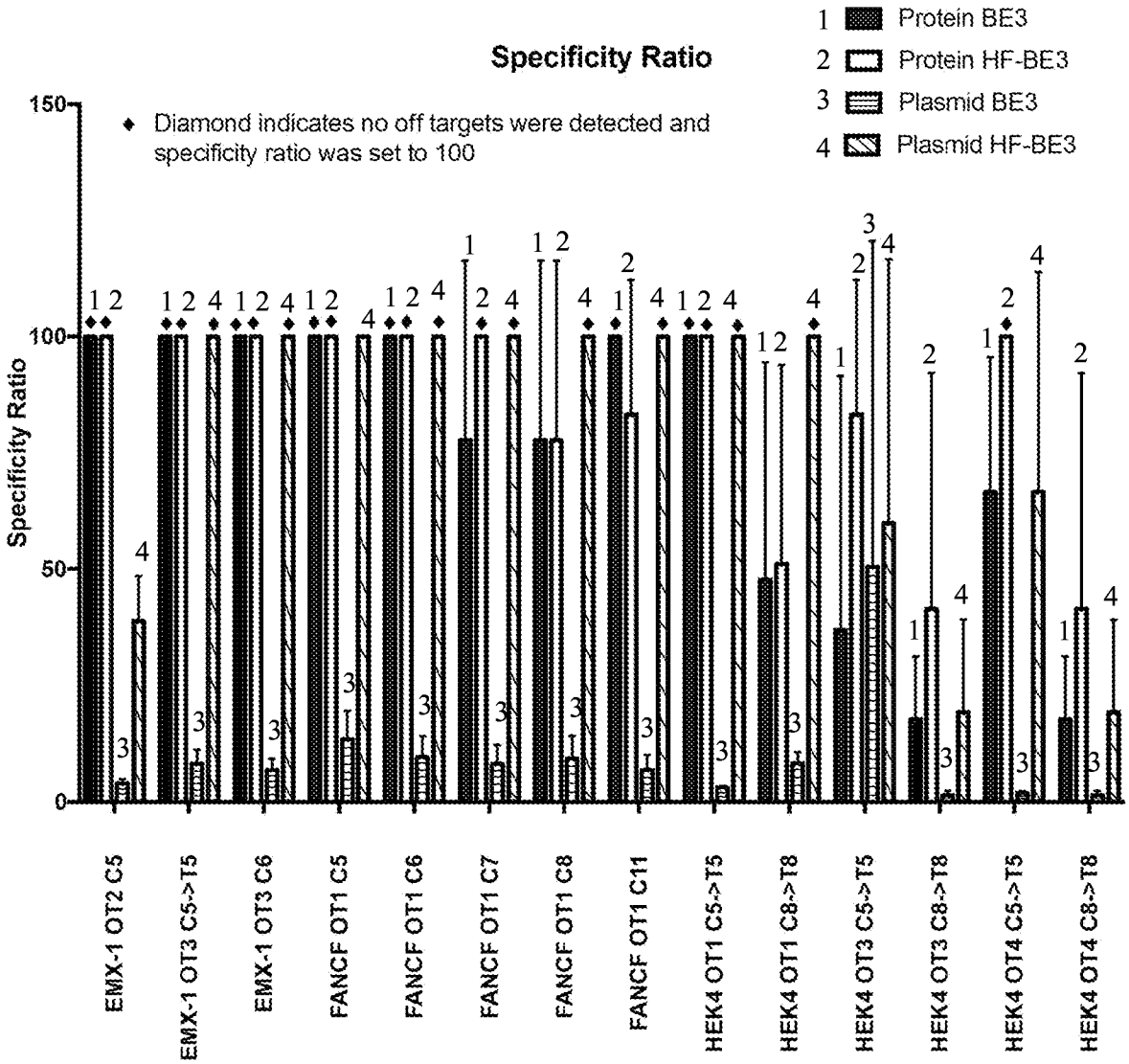

Specificity Ratio
1 Protein BE3
2 Protein HF-BE3
3 Plasmid BE3
4 Plasmid HF-BE3
◆ Diamond indicates no off targets were detected and specificity ratio was set to 100
Specificity Ratio
150
100
50
0
EMX-1 OT2 C5
EMX-1 OT3 C5->T5
EMX-1 OT3 C6
FANCF OT1 C5
FANCF OT1 C6
FANCF OT1 C7
FANCF OT1 C8
FANCF OT1 C11
HEK4 OT1 C5->T5
HEK4 OT1 C8->T8
HEK4 OT3 C5->T5
HEK4 OT3 C8->T8
HEK4 OT4 C5->T5
HEK4 OT4 C8->T8

FIGURE 102

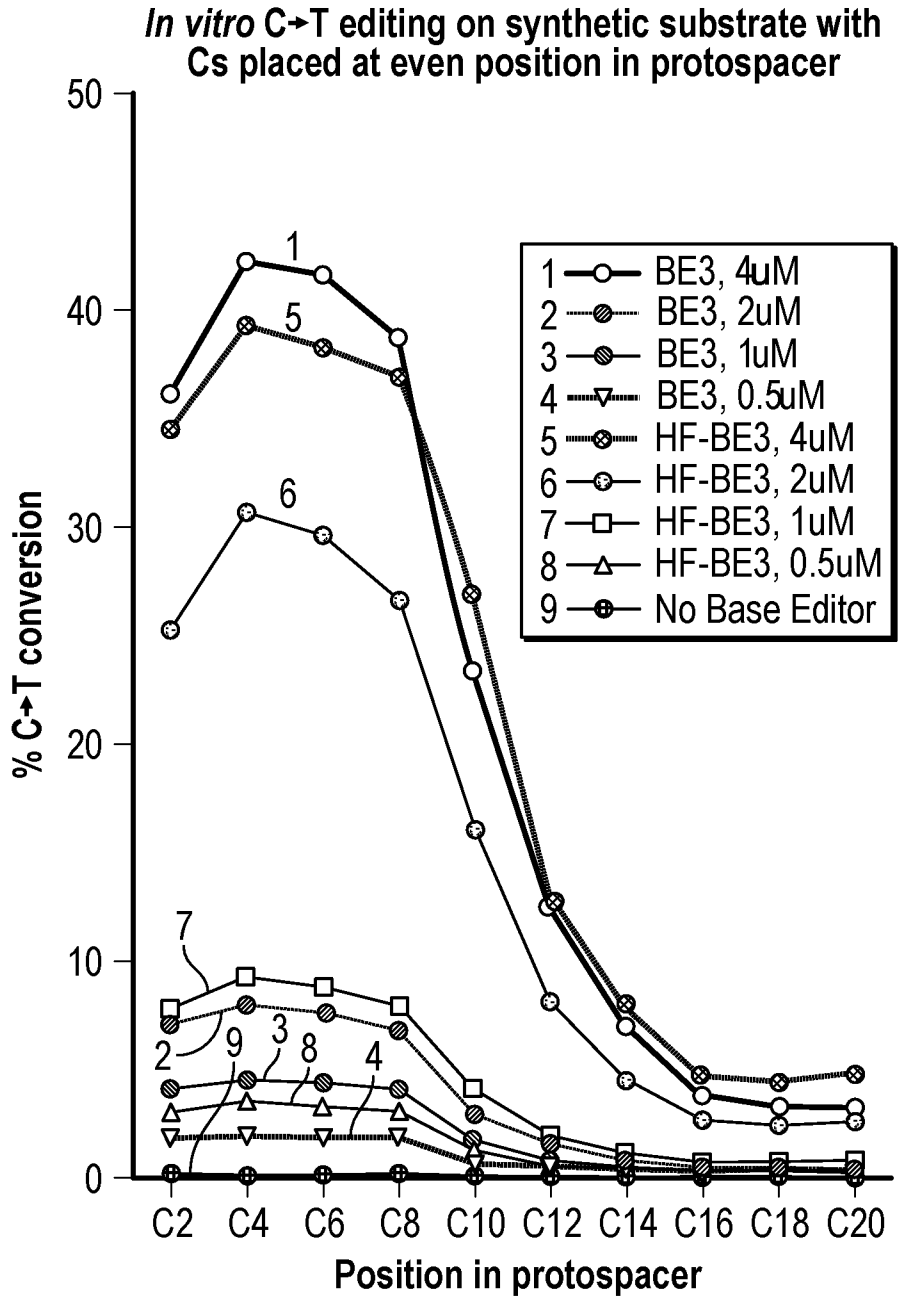

In vitro C→T editing on synthetic substrate with Cs placed at even position in protospacer
% C→T conversion
50
40
30
20
10
0
C2
C4
C6
C8
C10
C12
C14
C16
C18
C20
Position in protospacer
1 BE3, 4uM
2 BE3, 2uM
3 BE3, 1uM
4 BE3, 0.5uM
5 HF-BE3, 4uM
6 HF-BE3, 2uM
7 HF-BE3, 1uM
8 HF-BE3, 0.5uM
9 No Base Editor

FIGURE 103

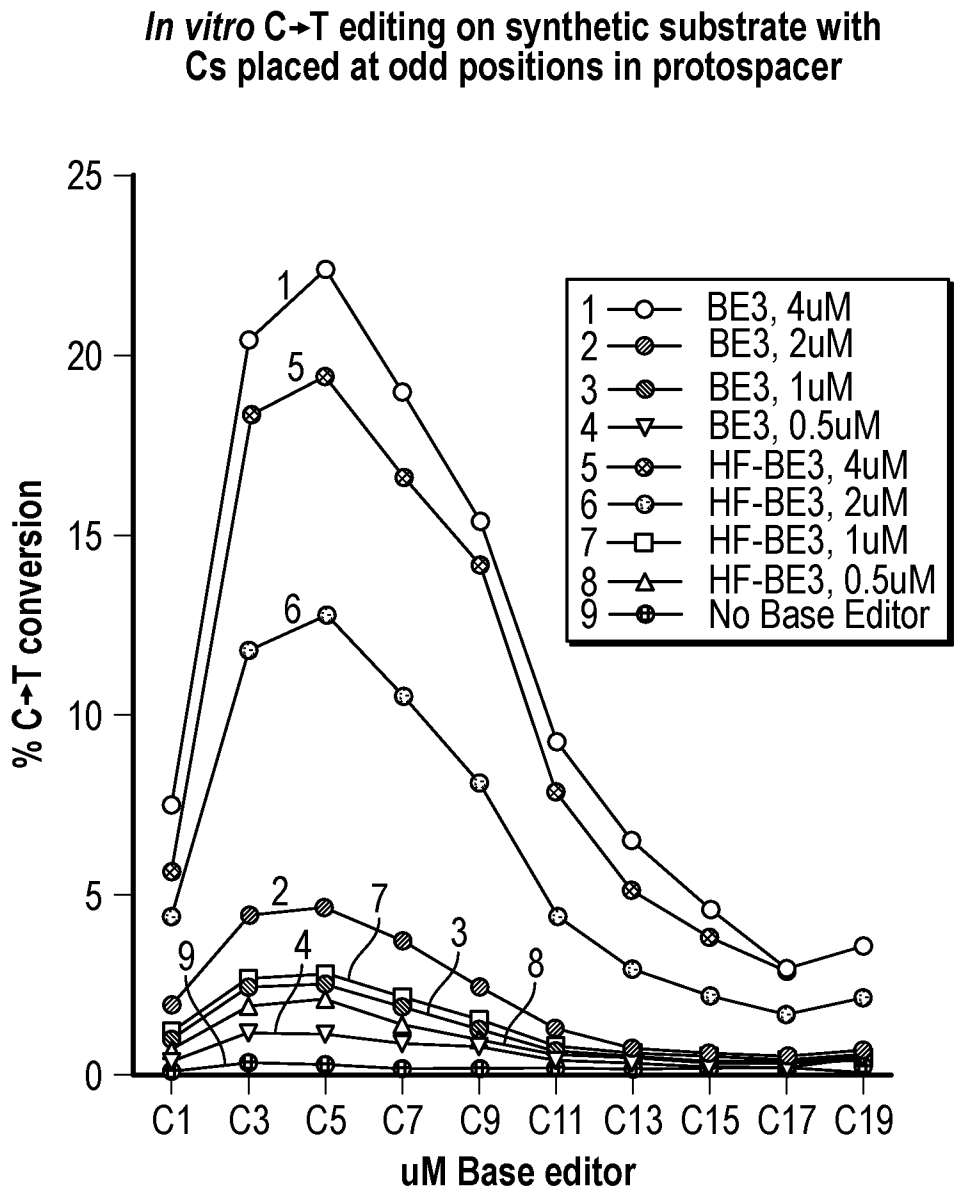

In vitro C→T editing on synthetic substrate with Cs placed at odd positions in protospacer
% C→T conversion
25
20
15
10
5
0
C1
C3
C5
C7
C9
C11
C13
C15
C17
C19
uM Base editor
1 BE3, 4uM
2 BE3, 2uM
3 BE3, 1uM
4 BE3, 0.5uM
5 HF-BE3, 4uM
6 HF-BE3, 2uM
7 HF-BE3, 1uM
8 HF-BE3, 0.5uM
9 No Base Editor

FIGURE 104

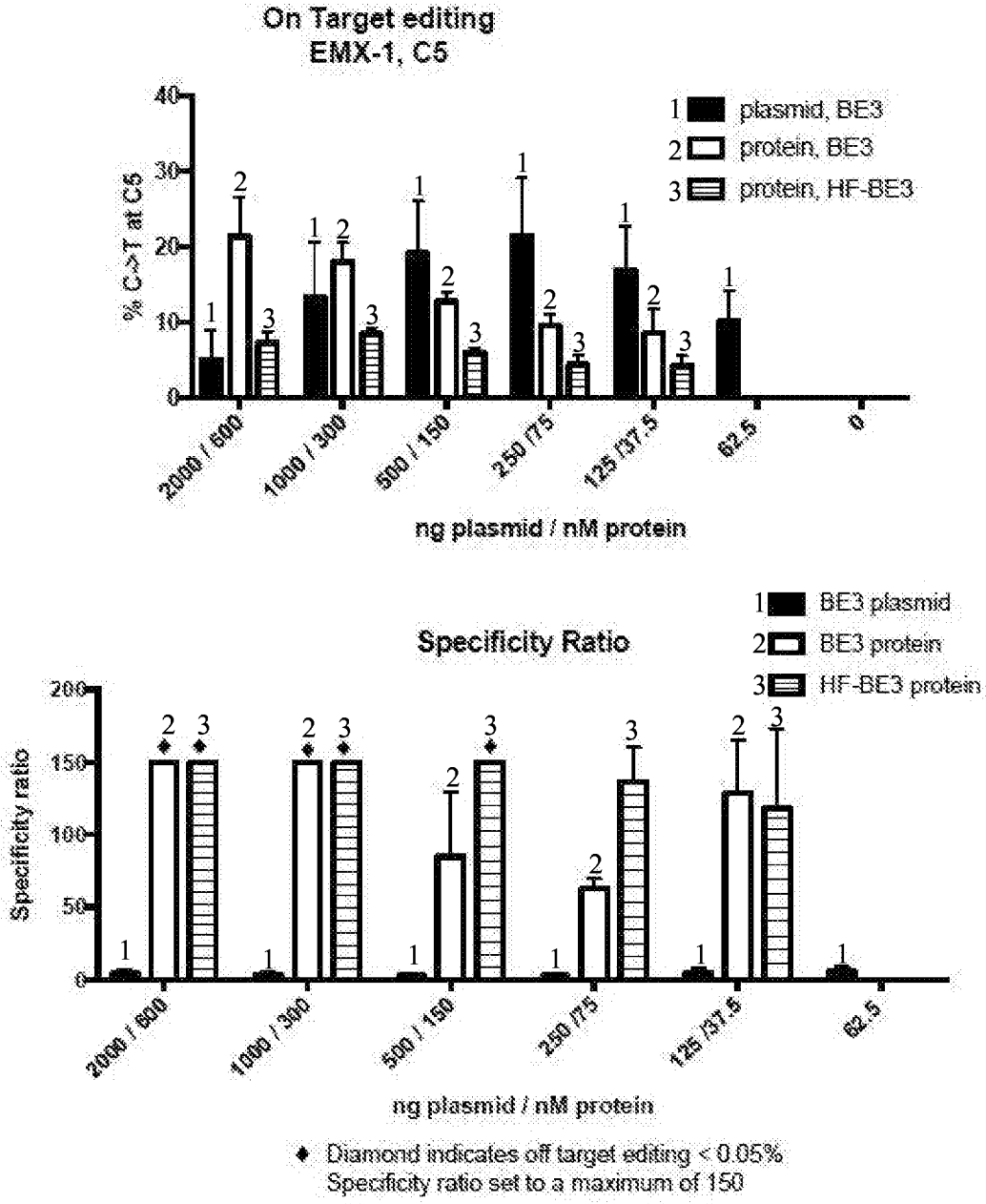

On Target editing
EMX-1, C5
% C->T at C5
1 plasmid, BE3
2 protein, BE3
3 protein, HF-BE3
2000 / 600
1000 / 300
500 / 150
250 /75
125 /37.5
62.5
0
ng plasmid / nM protein
Specificity Ratio
1 BE3 plasmid
2 BE3 protein
3 HF-BE3 protein
Specificity ratio
ng plasmid / nM protein
Diamond indicates off target editing < 0.05%
Specificity ratio set to a maximum of 150

FIGURE 105

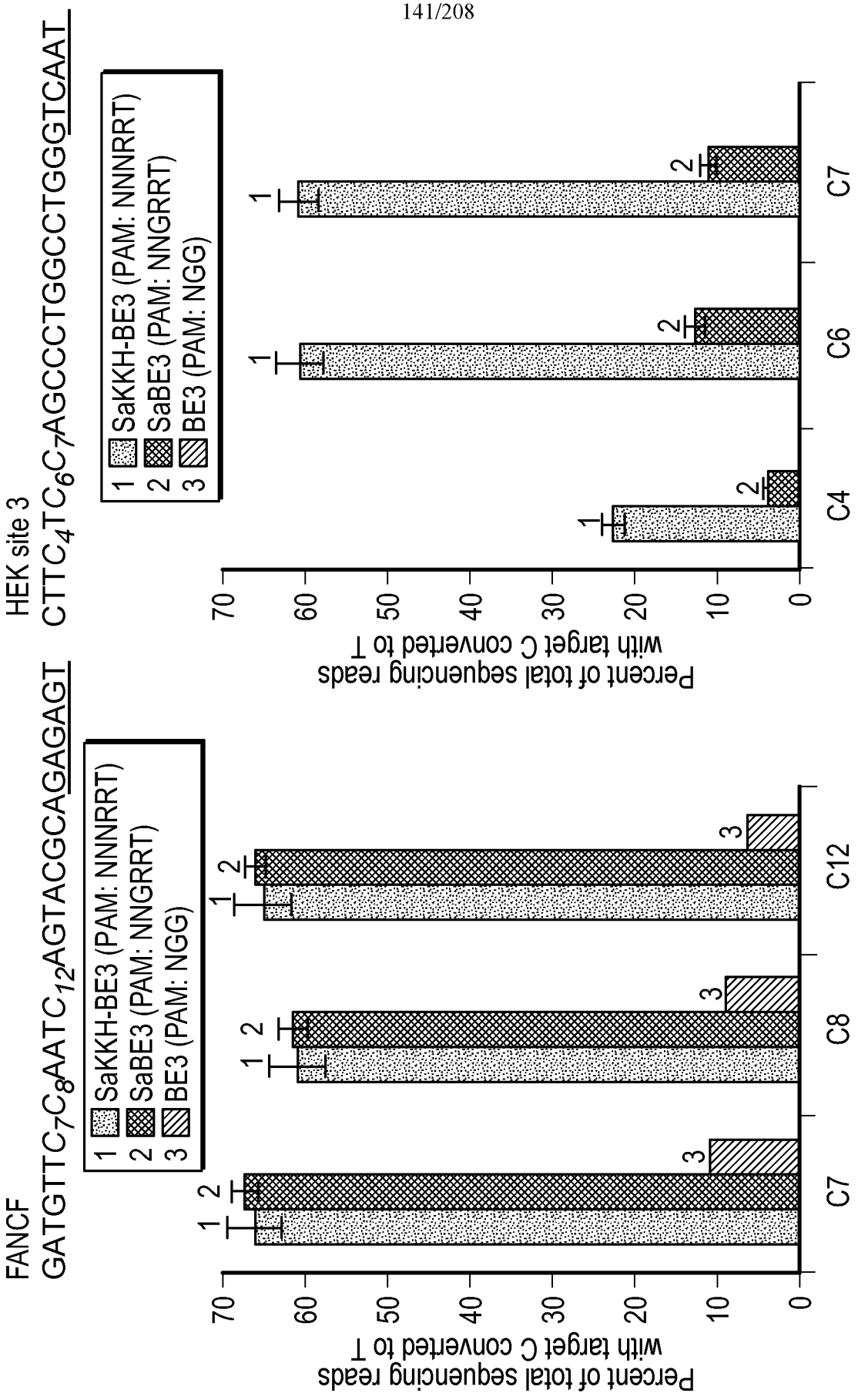

FANCF
GATGTTC7C8AATC12AGTACGCAGAGAGT
HEK site 3
CTTC4TC6C7AGCCCTGGCCTGGGTCAAT
1 SaKKH-BE3 (PAM: NNNRRT)
2 SaBE3 (PAM: NNGRRT)
3 BE3 (PAM: NGG)
Percent of total sequencing reads with target C converted to T
0
10
20
30
40
50
60
70
C4
C6
C7
C8
C12

FIGURE 106A

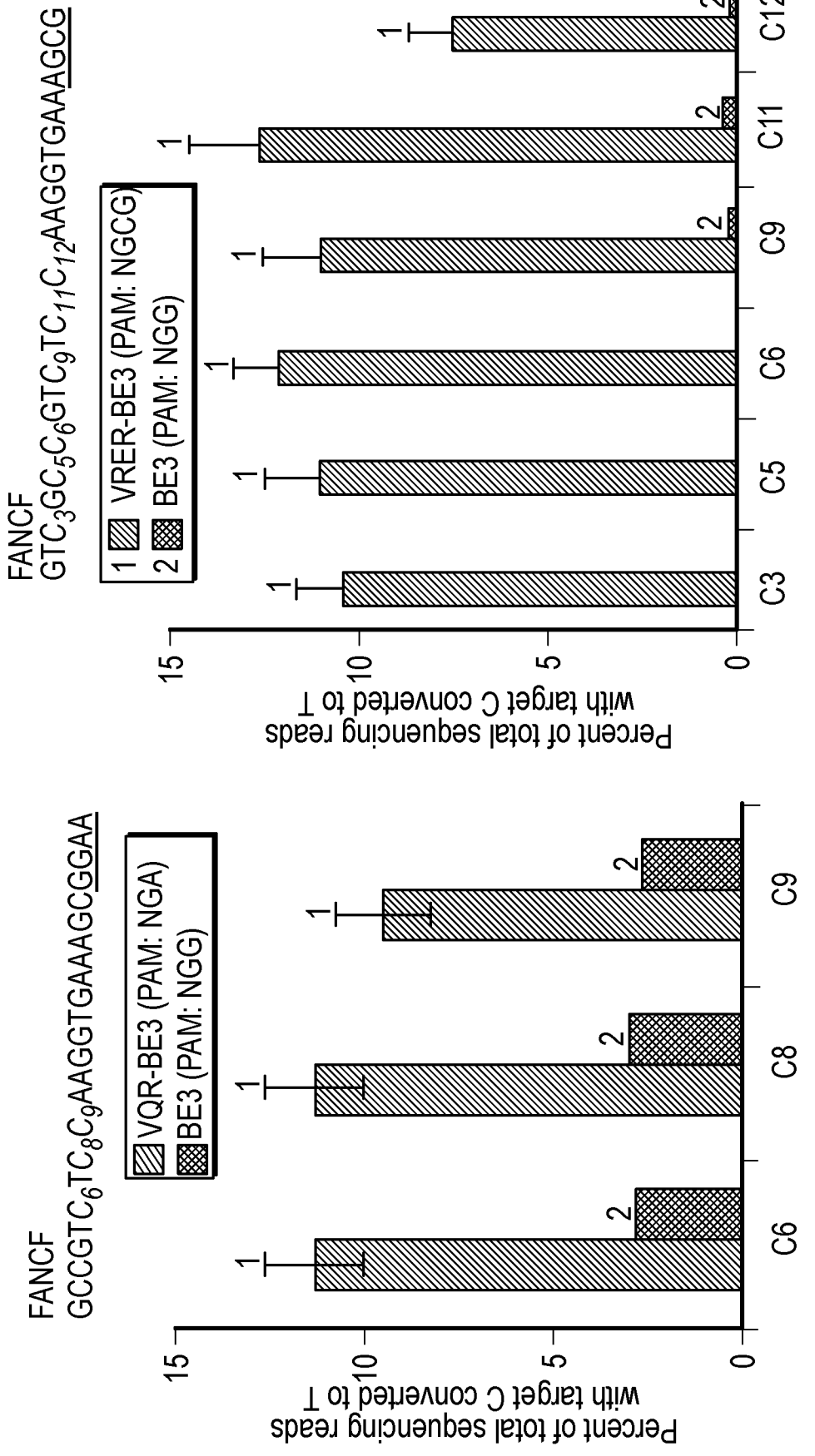

FANCF
GCCGTC6TC8C9AAGGTGAAAGCGGAA
VQR-BE3 (PAM: NGA)
BE3 (PAM: NGG)
Percent of total sequencing reads with target C converted to T
0
5
10
15
C6
C8
C9
FANCF
GTC3GC5C6GTC9TC11C12AAGGTGAAAGCG
1 VRER-BE3 (PAM: NGCG)
2 BE3 (PAM: NGG)
Percent of total sequencing reads with target C converted to T
0
5
10
15
C3
C5
C6
C9
C11
C12

FIGURE 106B

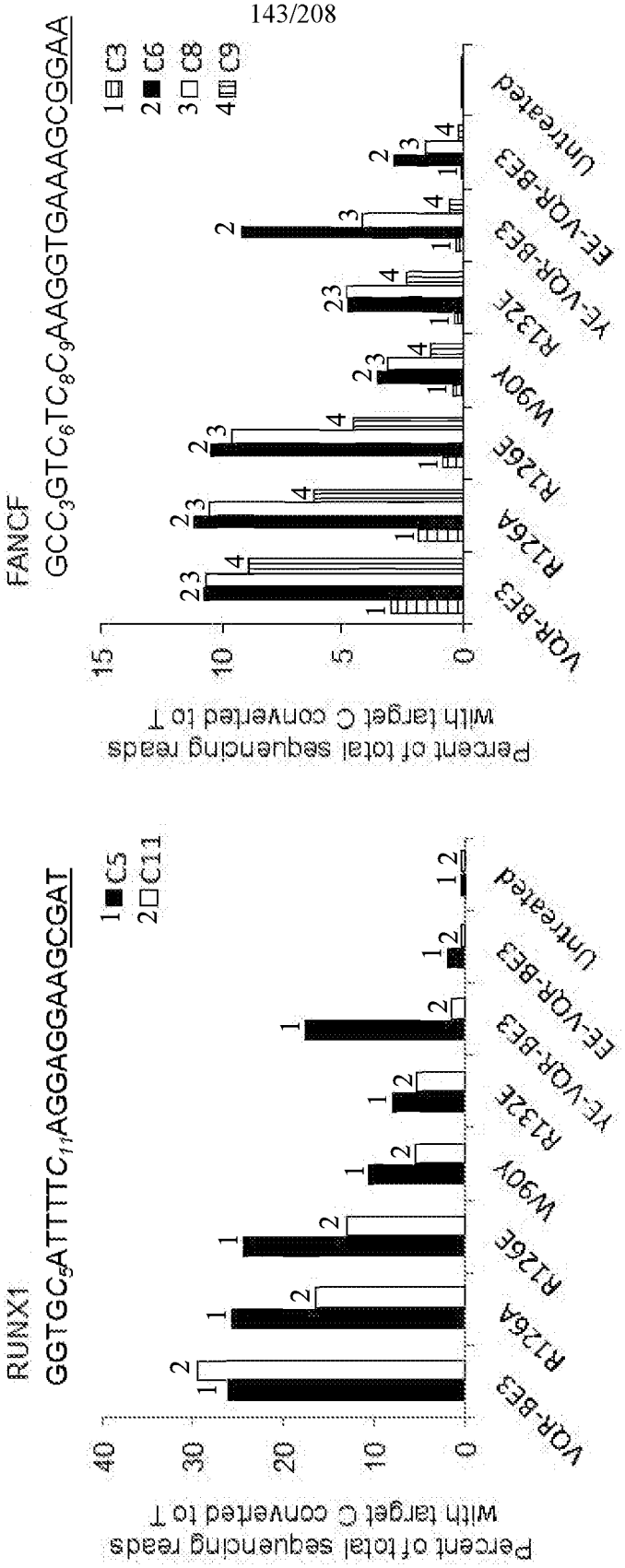

RUNX1
GGTGC5ATTTTC11AGGAGGAAGCGAT
1 C5
2 C11
Percent of total sequencing reads with target C converted to T
0
10
20
30
40
VQR-BE3
R126A
R126E
W90Y
R132E
VE-VQR-BE3
EE-VQR-BE3
Untreated
FANCF
GCC3GTC6TC8C9AAGGTGAAAGCGGAA
1 C3
2 C6
3 C8
4 C9
Percent of total sequencing reads with target C converted to T
0
5
10
15
VQR-BE3
R126A
R126E
W90Y
R132E
VE-VQR-BE3
EE-VQR-BE3
Untreated

FIGURE 107A

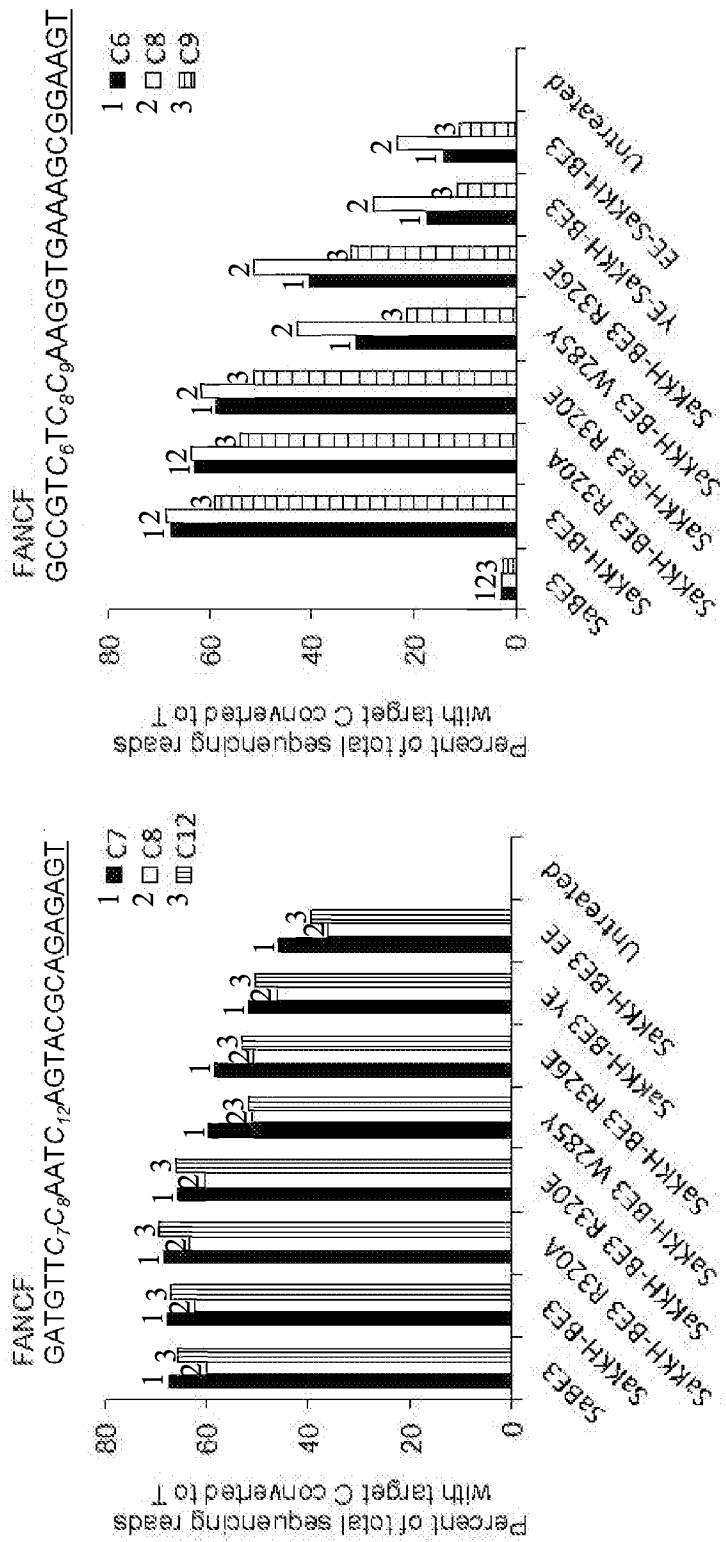

FANCF
GCCGTC6TC8C9AAGGTGAAAGCGGAAGT
1 C6
2 C8
3 C9
Percent of total sequencing reads with target C converted to T
0
20
40
60
80
saBE3
saKKH-BE3
saKKH-BE3 R320A
saKKH-BE3 R320E
saKKH-BE3 W285V
saKKH-BE3 R326E
saKKH-BE3 YE
saKKH-BE3 EE
Untreated
FANCF
GATGTTC7C8AATC12AGTACGCAGAGAGT
1 C7
2 C8
3 C12
Percent of total sequencing reads with target C converted to T
0
20
40
60
80
saBE3
saKKH-BE3
saKKH-BE3 R320A
saKKH-BE3 R320E
saKKH-BE3 W285V
saKKH-BE3 R326E
saKKH-BE3 YE
saKKH-BE3 EE
Untreated

FIGURE 107B

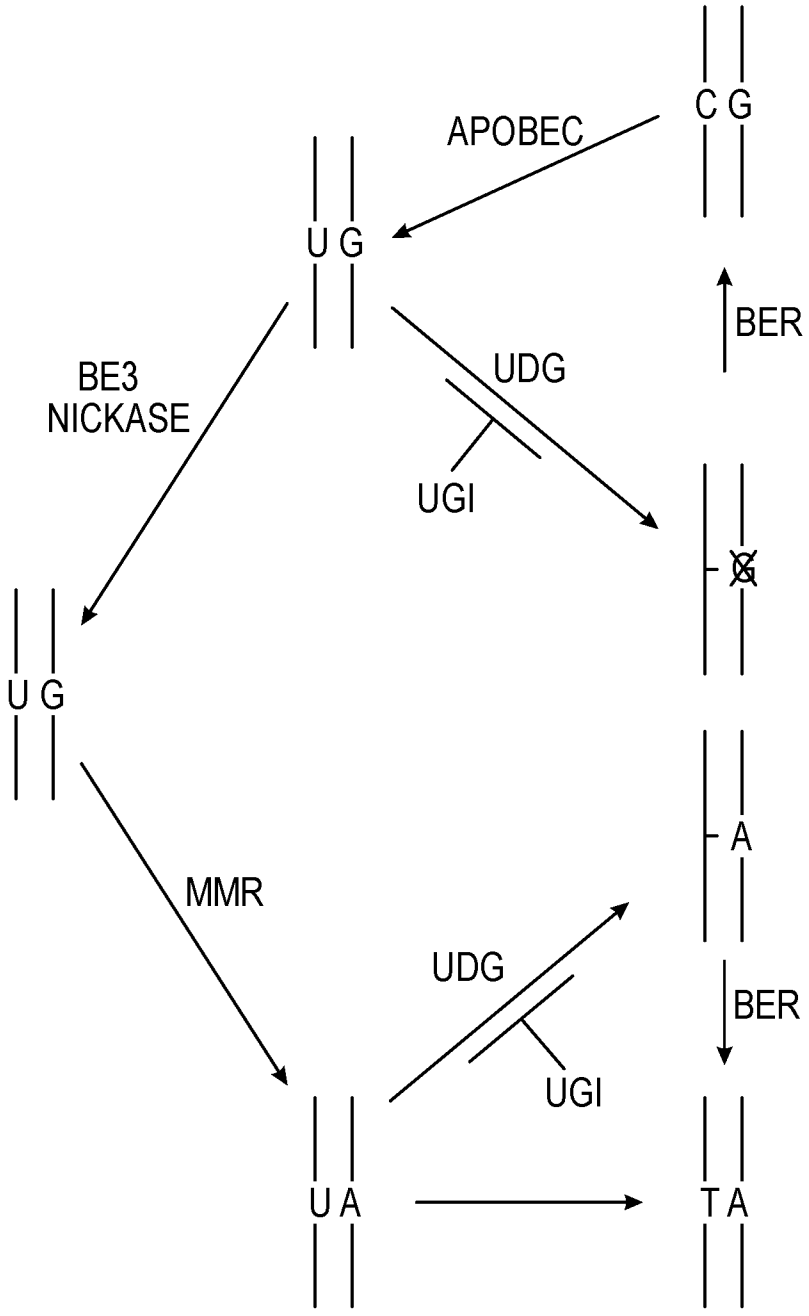


FIGURE 108

BE3:
rAPOBEC1
XTEN
dCas9(A840H)
SGGS
UGI
BE4-pmCDA1:
pmCDA1
XTEN
dCas9(A840H)
SGGS
UGI
BE4-hAID:
hAID
XTEN
dCas9(A840H)
SGGS
UGI
BE4-3G:
hAPOBEC3G
XTEN
dCas9(A840H)
SGGS
UGI
BE4-N:
UGI
(GGS)3
rAPOBEC1
XTEN
dCas9(A840H)
BE4-SSB:
rAPOBEC1
XTEN
dCas9(A840H)
(GGS)3
SSB
BE4-(GGS)3:
rAPOBEC1
XTEN
dCas9(A840H)
(GGS)3
UGI
BE4-XTEN:
rAPOBEC1
XTEN
dCas9(A840H)
XTEN
UGI
BE4-32aa:
rAPOBEC1
32aa
dCas9(A840H)
SGGS
UGI
BE4-2XUGI:
rAPOBEC1
XTEN
dCas9(A840H)
SGGS
UGI
SGGS
UGI
BE4:
rAPOBEC1
32aa
dCas9(A840H)
(GGS)3
UGI
(GGS)3
UGI
EMX1:
GAGTCCGAGCAGAAGAAGAAGGG
FANCF:
GGAATCCCTTCTGCAGCACCTGG
HEK2:
GAACACAAAGCATAGACTGCGGG
HEK3:
GGCCCAGACTGAGCACGTGATGG
HEK4:
GGCACTGCGGCTGGAGGTCCGGG
RNF2:
GTCATCTTAGTCATTACCTGAGG

FIGURE 109

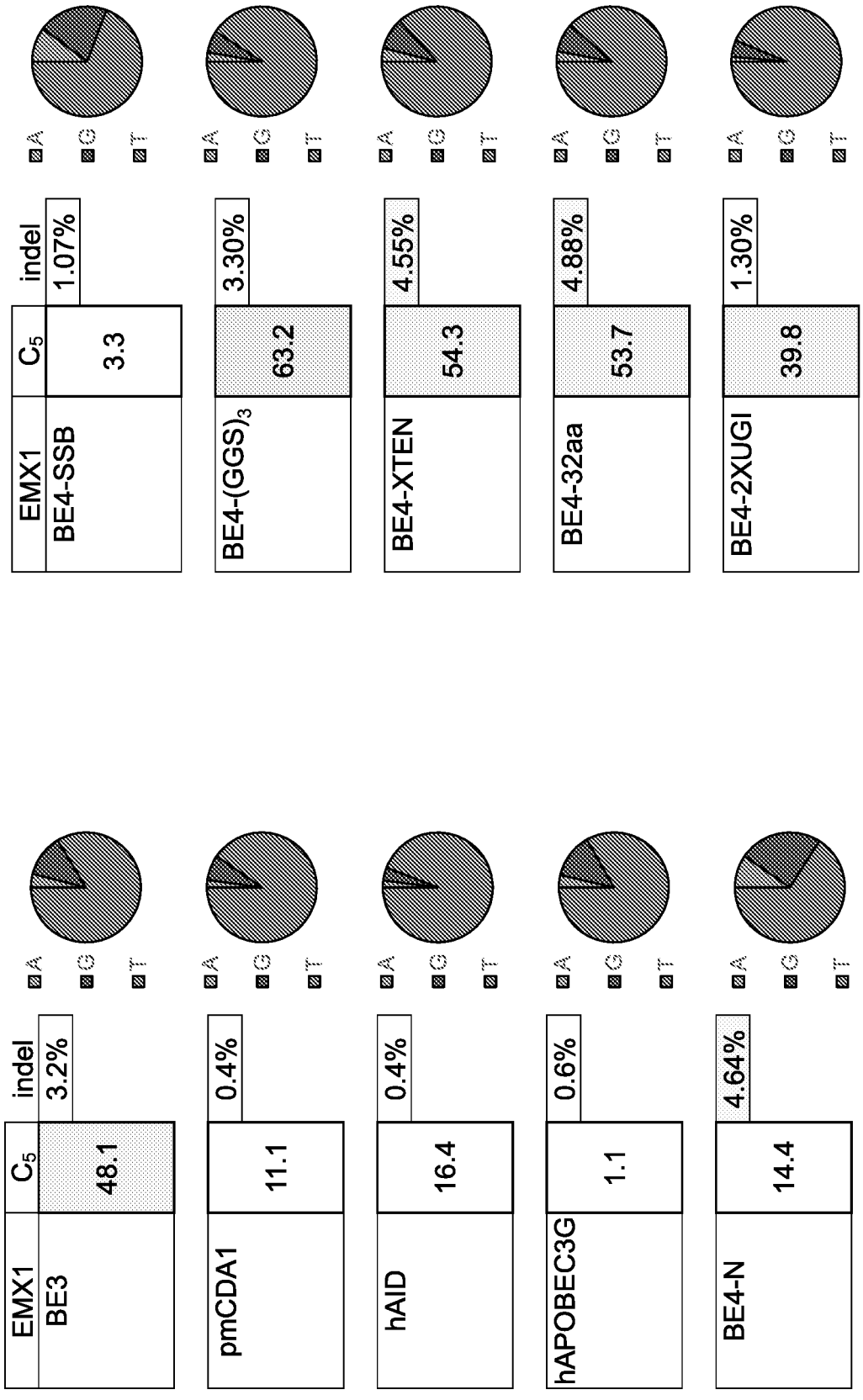

EMX1
BE3
C5
48.1
indel
3.2%
A
G
T
pmCDA1
11.1
0.4%
hAID
16.4
0.4%
hAPOBEC3G
1.1
0.6%
BE4-N
14.4
4.64%
BE4-SSB
3.3
1.07%
BE4-(GGS)3
63.2
3.30%
BE4-XTEN
54.3
4.55%
BE4-32aa
53.7
4.88%
BE4-2XUGI
39.8
1.30%

FIGURE 110

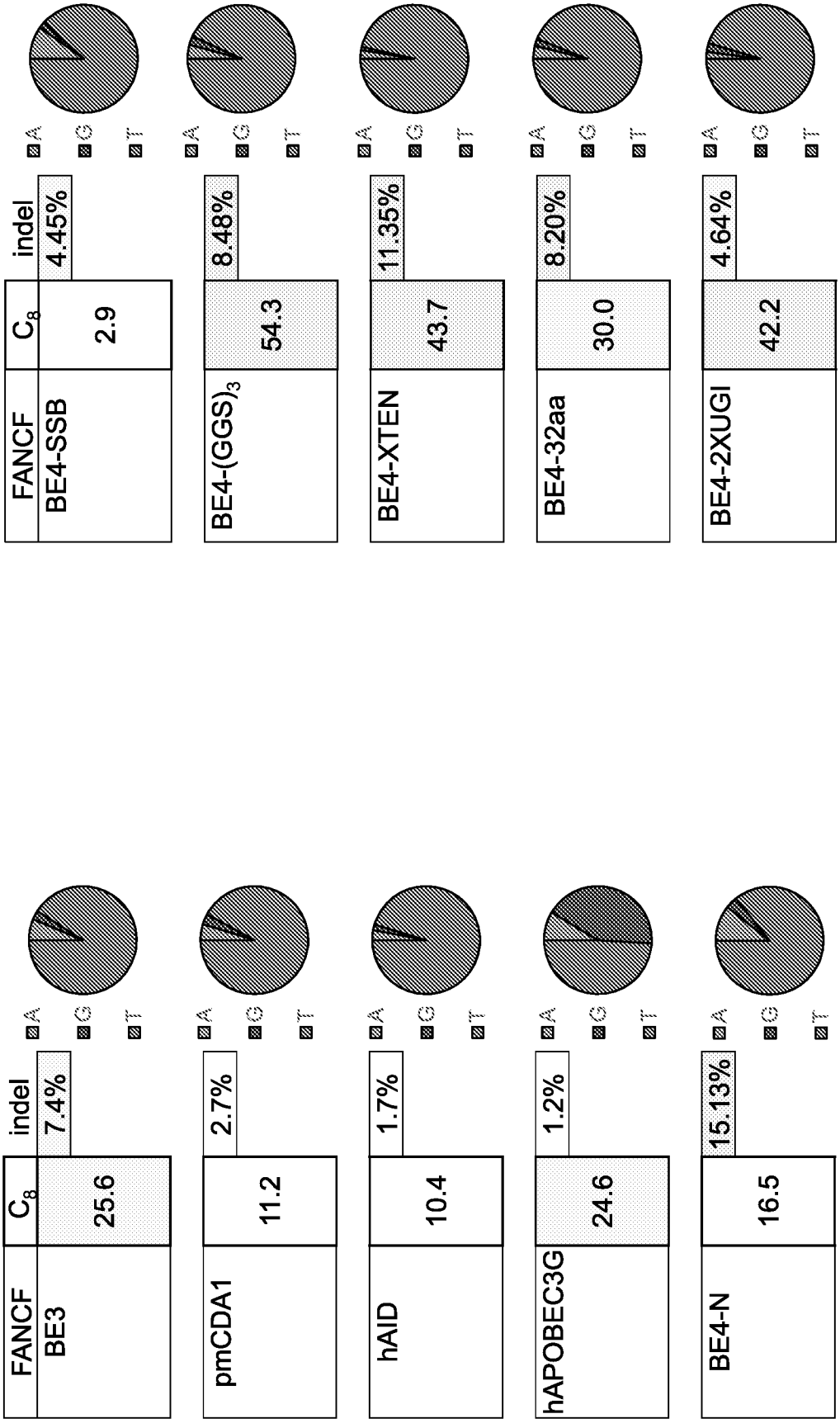
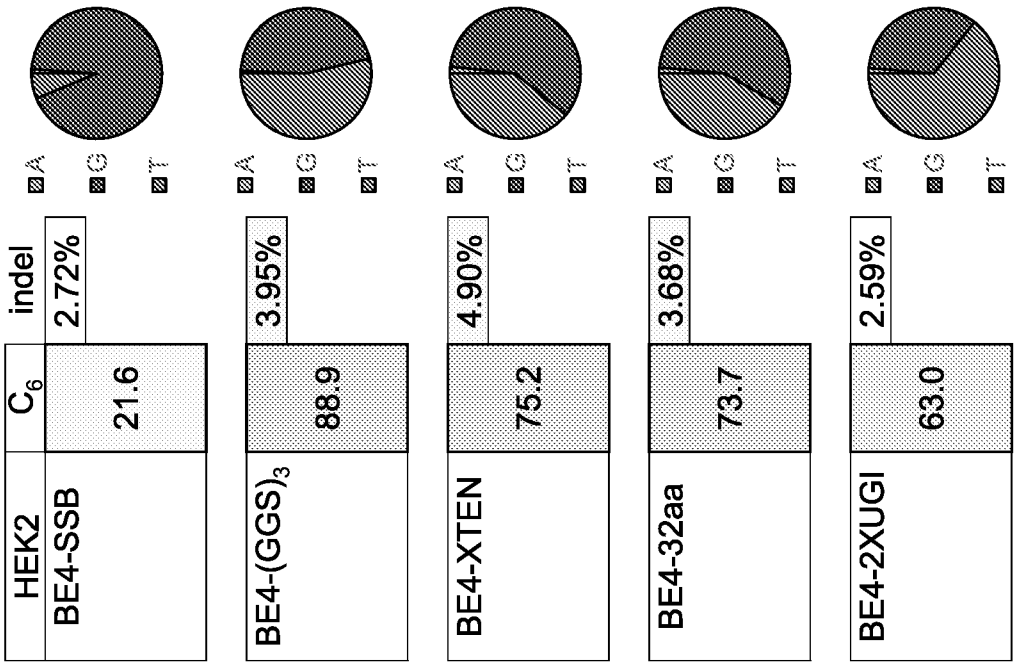

FANCF
BE4-SSB
C8
2.9
indel
4.45%
A
G
T
BE4-(GGS)3
54.3
8.48%
BE4-XTEN
43.7
11.35%
BE4-32aa
30.0
8.20%
BE4-2XUGI
42.2
4.64%
BE3
25.6
7.4%
pmCDA1
11.2
2.7%
hAID
10.4
1.7%
hAPOBEC3G
24.6
1.2%
BE4-N
16.5
15.13%

FIGURE 111


HEK2
BE4-SSB
C6
21.6
indel
2.72%
A
G
T
BE4-(GGS)3
88.9
3.95%
BE4-XTEN
75.2
4.90%
BE4-32aa
73.7
3.68%
BE4-2XUGI
63.0
2.59%

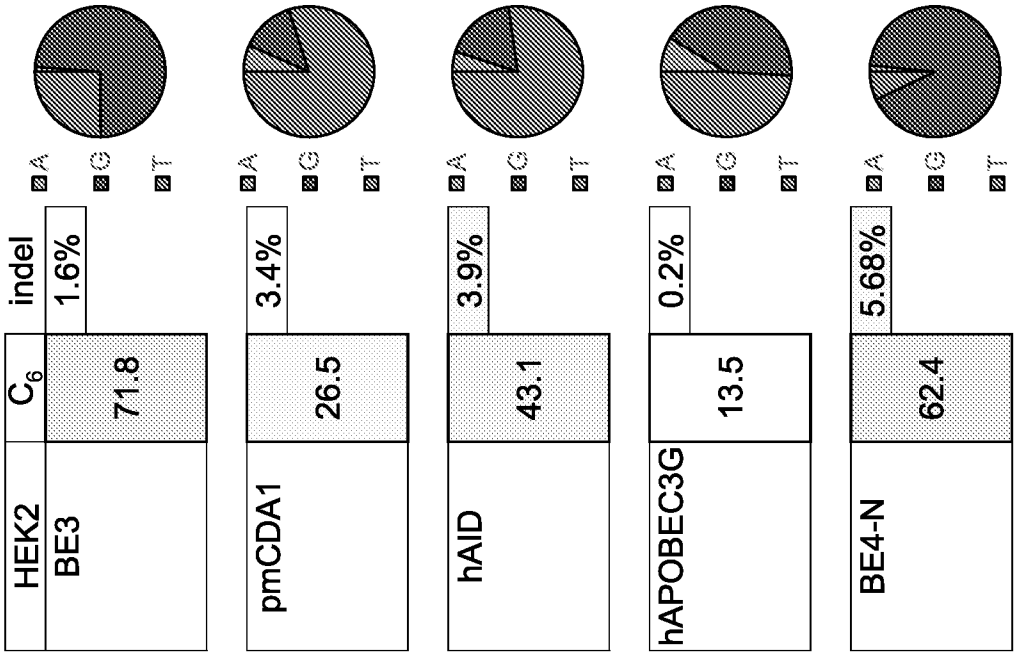

HEK2
BE3
C6
71.8
indel
1.6%
A
G
T
pmCDA1
26.5
3.4%
hAID
43.1
3.9%
hAPOBEC3G
13.5
0.2%
BE4-N
62.4
5.68%

FIGURE 112

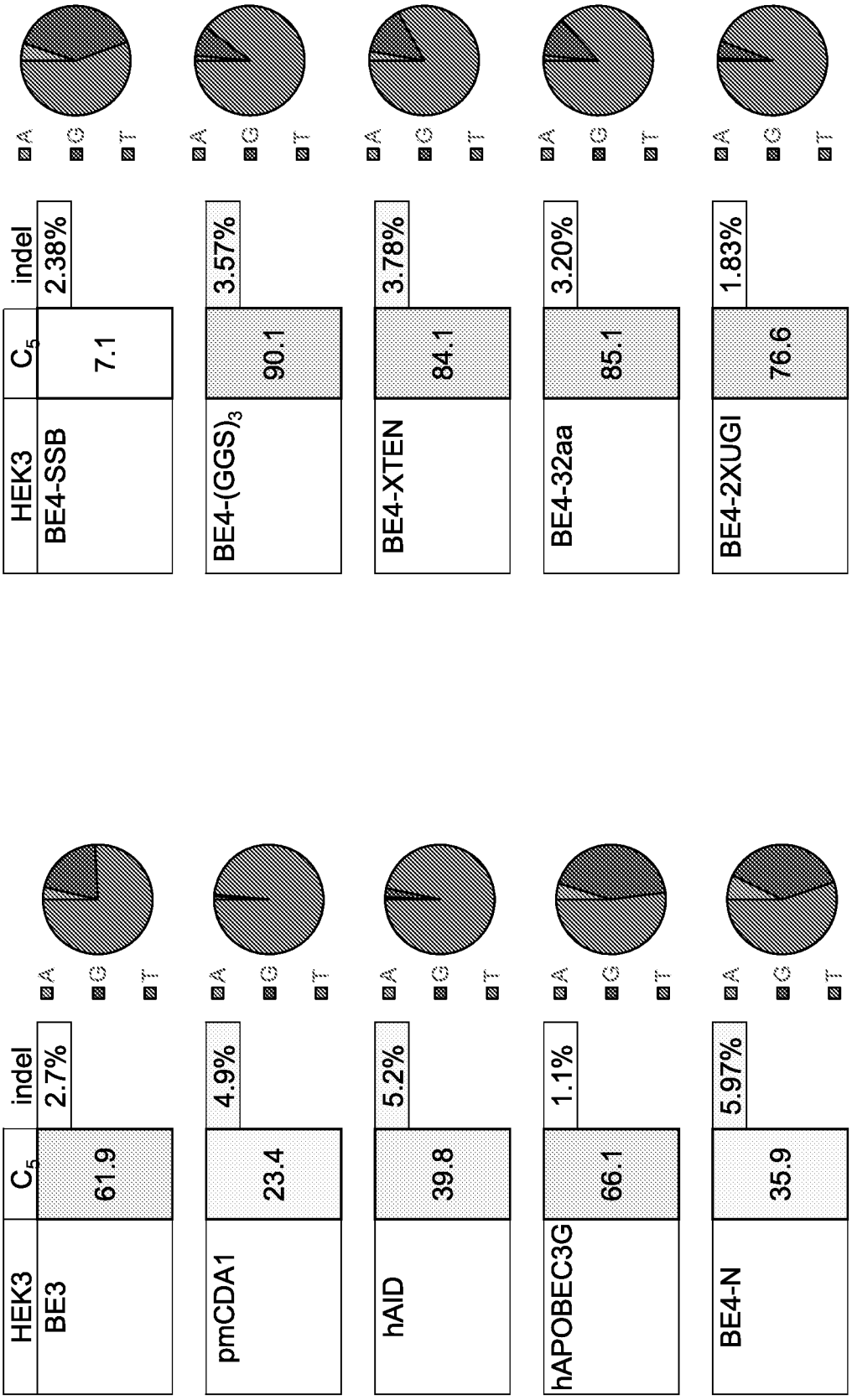
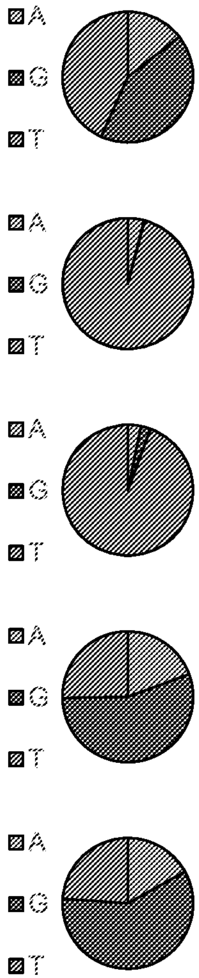

HEK3
BE4-SSB
C5
7.1
indel
2.38%
A
G
T
BE4-(GGS)3
90.1
3.57%
BE4-XTEN
84.1
3.78%
BE4-32aa
85.1
3.20%
BE4-2XUGI
76.6
1.83%
BE3
61.9
2.7%
pmCDA1
23.4
4.9%
hAID
39.8
5.2%
hAPOBEC3G
66.1
1.1%
BE4-N
35.9
5.97%

FIGURE 113

HEK4	C_5	indel
BE3	58.1	4.0%
pmCDA1	12.5	2.5%
hAID	28.7	5.8%
hAPOBEC3G	0.6	0.2%
BE4-N	43.4	6.13%



HEK4	C_5	indel
BE4-SSB	13.2	5.38%
BE4-(GGS)$_3$	87.0	4.22%
BE4-XTEN	80.3	6.24%
BE4-32aa	78.2	6.78%
BE4-2XUGI	77.8	2.23%

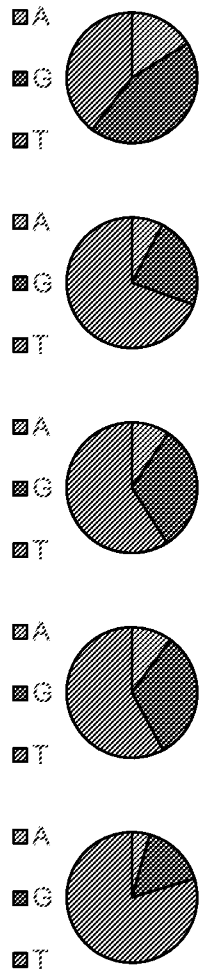


FIGURE 114

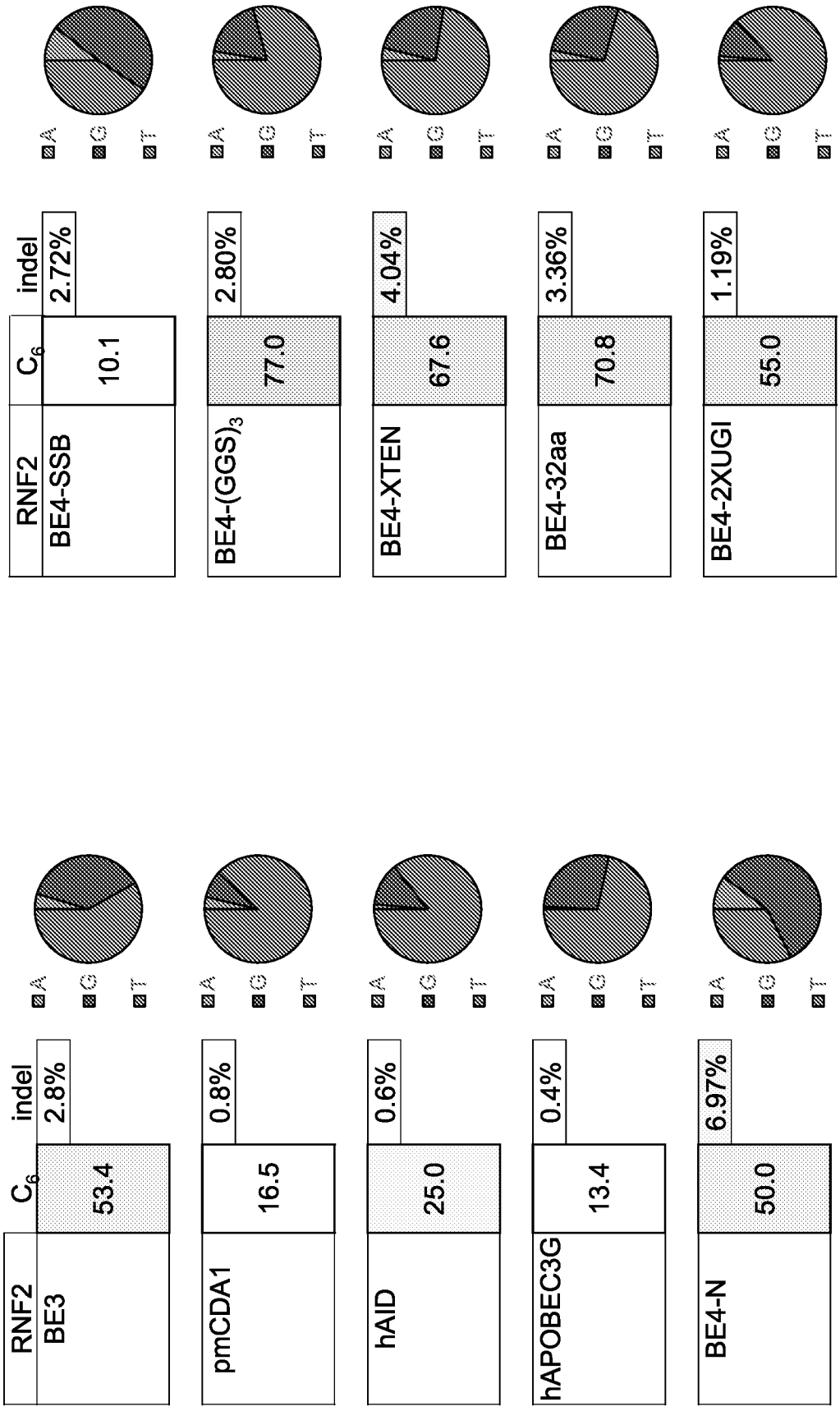
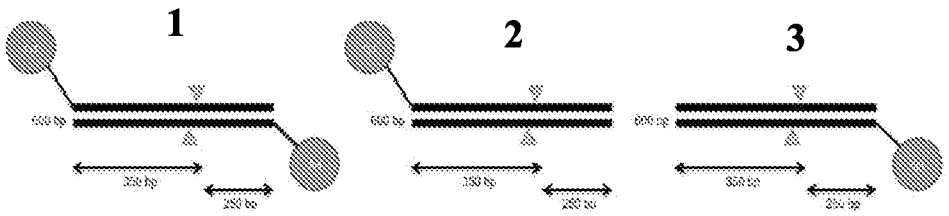

RNF2
BE3
C6
53.4
indel
2.8%
pmCDA1
16.5
0.8%
hAID
25.0
0.6%
hAPOBEC3G
13.4
0.4%
BE4-N
50.0
6.97%
BE4-SSB
10.1
2.72%
BE4-(GGS)3
77.0
2.80%
BE4-XTEN
67.6
4.04%
BE4-32aa
70.8
3.36%
BE4-2XUGI
55.0
1.19%
A
G
T

FIGURE 115



Top strand: non-target strand; Bottom strand: target strand

FIGURE 116

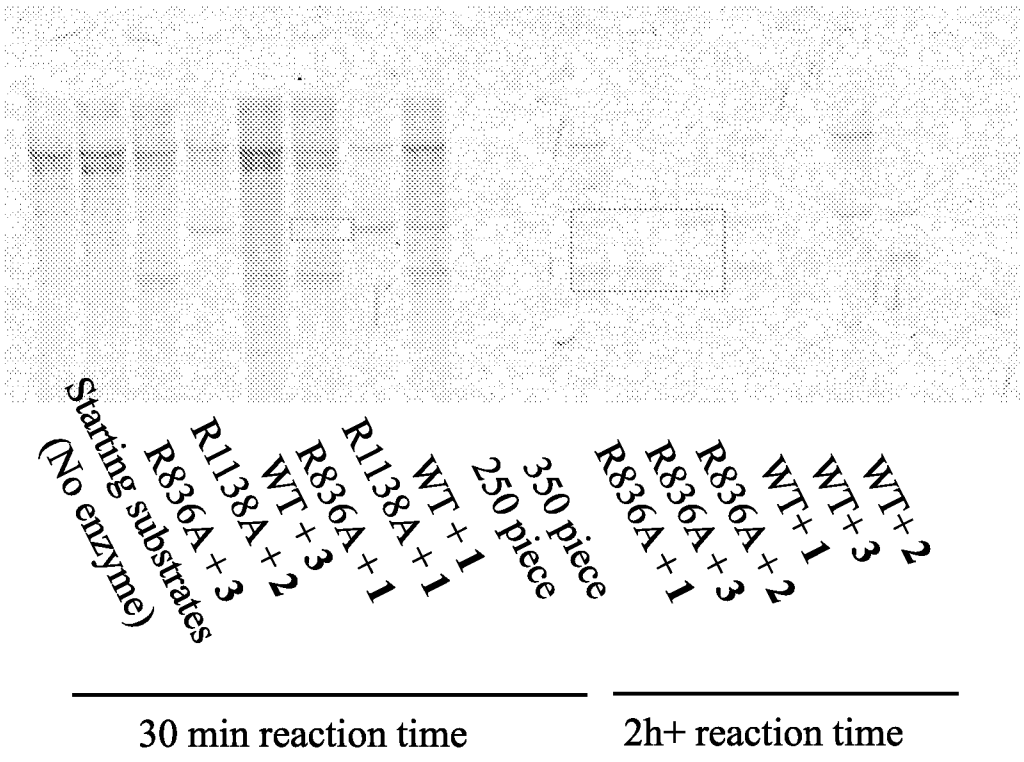


FIGURE 117

PAM

FANCF1	TTTC	G1	C2	G3	G4	G7	C10	C11	C15	G17	C20	C22
BG		0.01005	0.03022	0.01083	0.00933	0.00502	0.01872	0.02014	0.03740	0.00504	0.01938	0.02099
AsCpf1		0.00895	0.02769	0.07439	0.00672	0.06495	0.02395	0.01795	0.02691	0.01570	0.02237	0.02406
APOBEC-AsCpf1(R912A)-UGI		0.00883	0.02606	0.01430	0.00539	0.00490	0.29513	0.28158	0.04373	0.01425	0.02795	0.01631
APOBEC-AsCpf1(R1225A)-UGI		0.00796	0.03754	0.01131	0.00749	0.01124	0.02769	0.02908	0.03346	0.02016	0.01732	0.01651

FANCF2	TTTC	C1	G2	G3	C4	C7	G9	G10	C11	G12	G13	C15	C16	G17	C18
BG		0.02628	0.01963	0.02346	0.00318	0.00798	0.00844	0.00850	0.08758	0.05168	0.00637	0.00499	0.07694	0.01382	0.01805
AsCpf1		0.03489	0.03325	0.05248	0.00392	0.00722	0.02484	0.00791	0.06139	0.10805	0.01851	0.02139	0.10988	0.00927	0.00992
APOBEC-AsCpf1(R912A)-UGI		0.02440	0.02677	0.00699	0.00481	0.19399	0.01115	0.00534	0.11463	0.01660	0.00427	0.00722	0.10175	0.01119	0.01065
APOBEC-AsCpf1(R1225A)-UGI		0.02855	0.04216	0.00322	0.00479	0.00587	0.01004	0.00585	0.06634	0.00586	0.00638	0.00444	0.07056	0.00797	0.01061

HEK3-3	TTTC	C3	C6	C8	C9	C12	C13	C14	C18	C19
BG		0.00907	0.00726	0.01440	0.01539	0.01084	0.01181	0.00729	0.01540	0.00634
AsCpf1		0.01118	0.00670	0.01994	0.00892	0.01334	0.01342	0.01797	0.01118	0.03351
APOBEC-AsCpf1(R912A)-UGI		0.00647	0.02264	0.44917	0.30191	0.03754	0.12192	0.01513	0.01291	0.00754
APOBEC-AsCpf1(R1225A)-UGI		0.00770	0.00769	0.10243	0.07236	0.01202	0.00770	0.00551	0.01365	0.00548

PAM

HEK3-4		G23	C20	G18	G16	C15	C13	G12	G10	G7	G6	C5	G1	GAAA
BG		0.00638	0.01001	0.03459	0.01008	0.01008	0.01285	0.01284	0.00548	0.01458	0.02557	0.01099	0.00636	
AsCpf1		0.00988	0.00862	0.02466	0.01984	0.03343	0.02600	0.00865	0.00617	0.01718	0.02216	0.01232	0.00488	
APOBEC-AsCpf1(R912A)-UGI		0.00323	0.01185	0.03012	0.00759	0.01735	0.01734	0.04550	0.15878	0.00862	0.02266	0.00542	0.00644	
APOBEC-AsCpf1(R1225A)-UGI		0.01142	0.01045	0.03044	0.01435	0.01626	0.02104	0.01051	0.01238	0.01045	0.03428	0.01810	0.00473	

BG –background; R912A – B E 3 ; R1225A –self-defeating BE (nicks non-target strand)

FIGURE 118

FIGURE 119

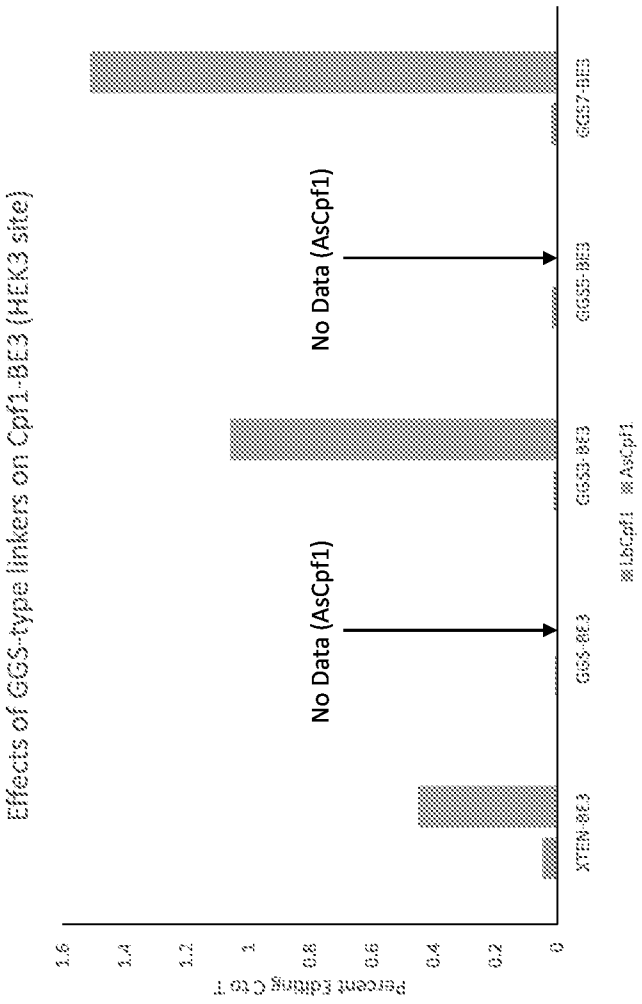


FIGURE 120

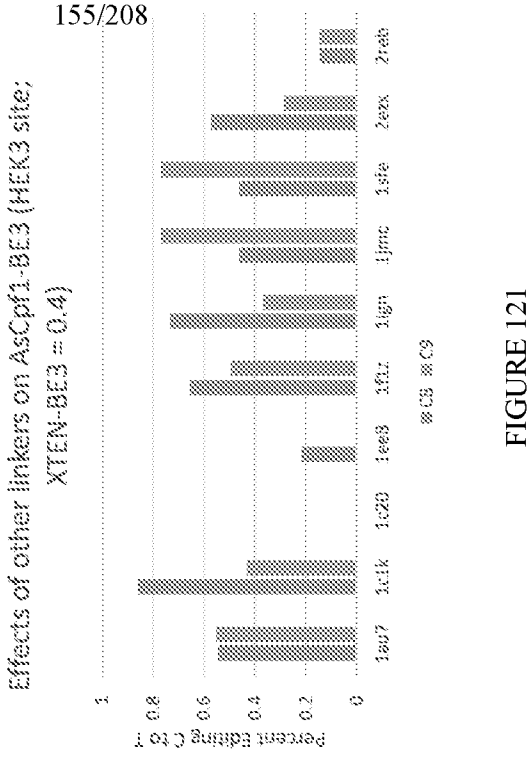


FIGURE 121

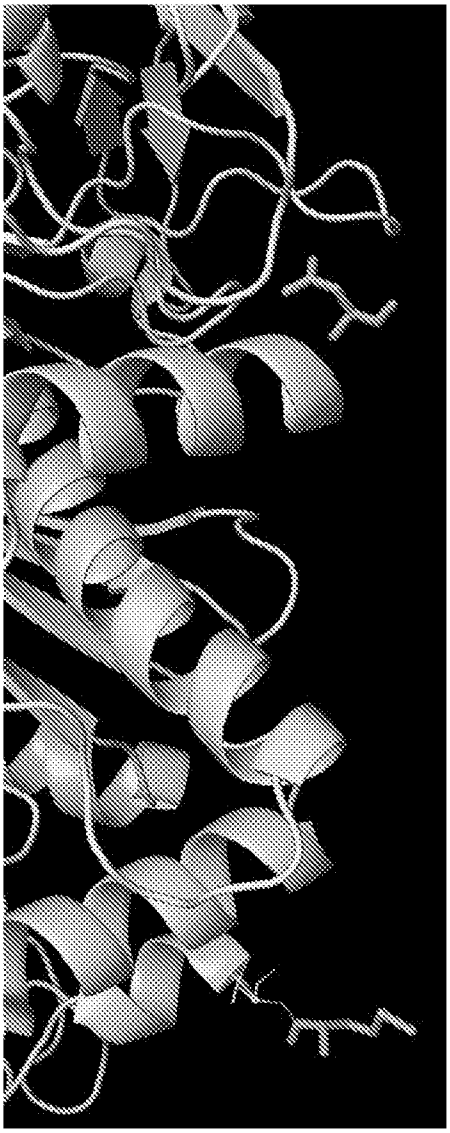


FIGURE 122

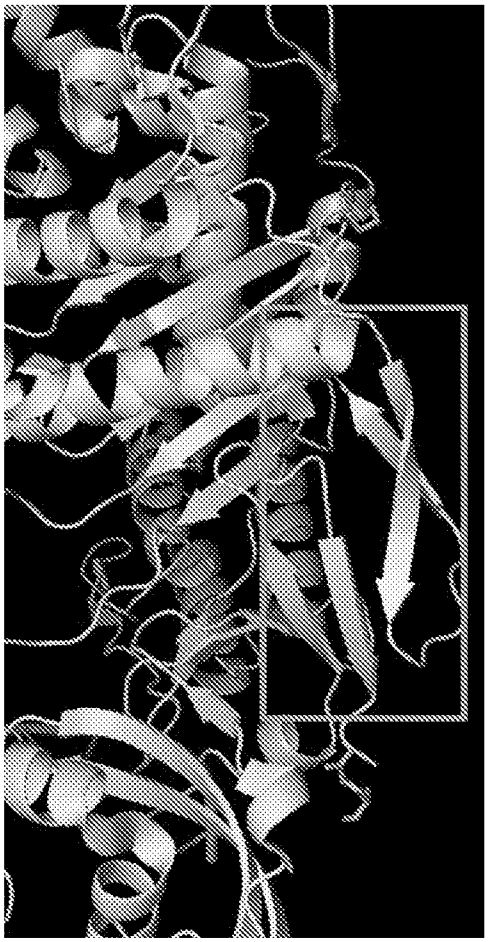


FIGURE 123

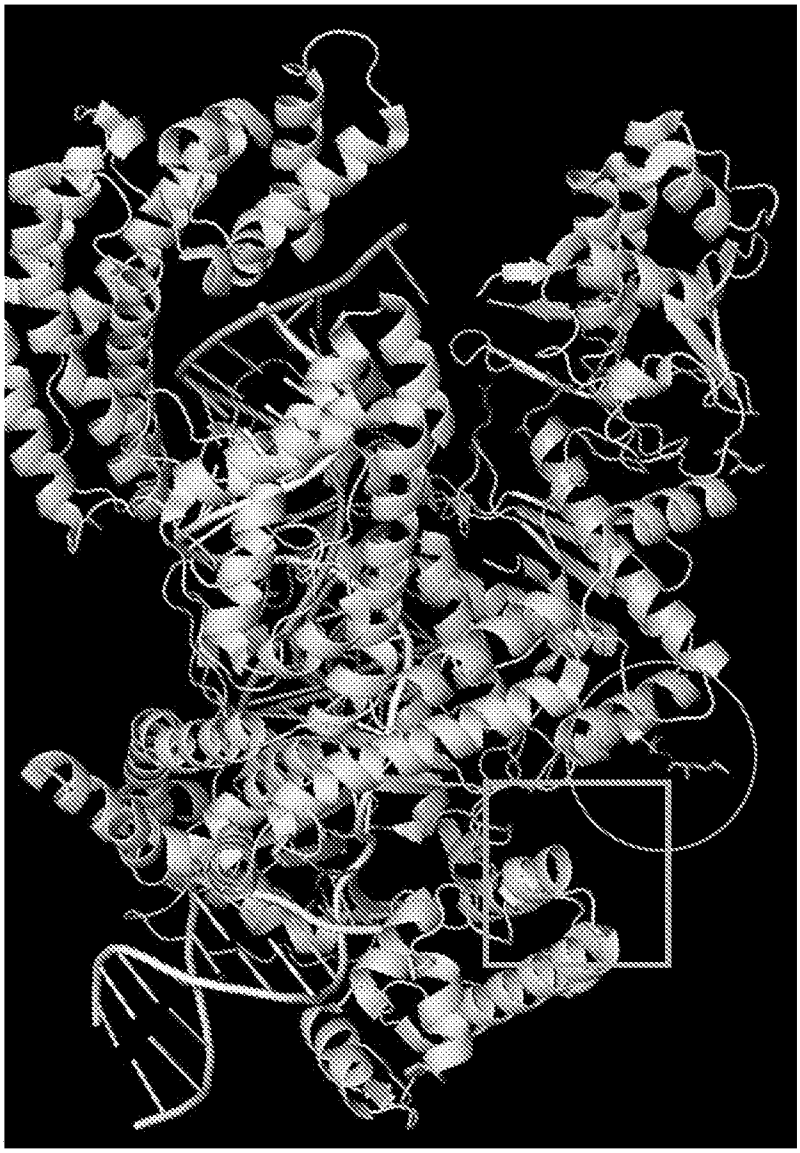


FIGURE 124

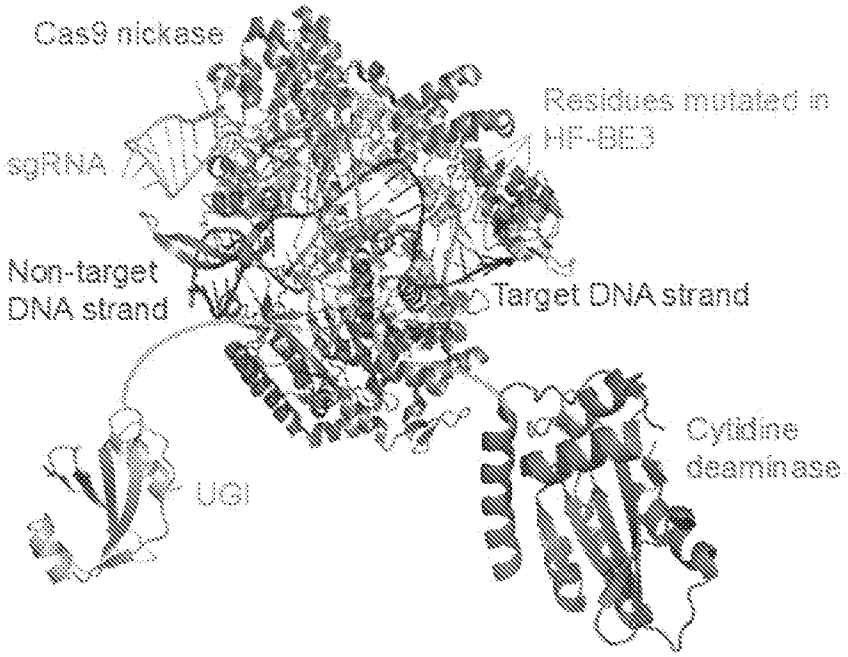


FIGURE 125A

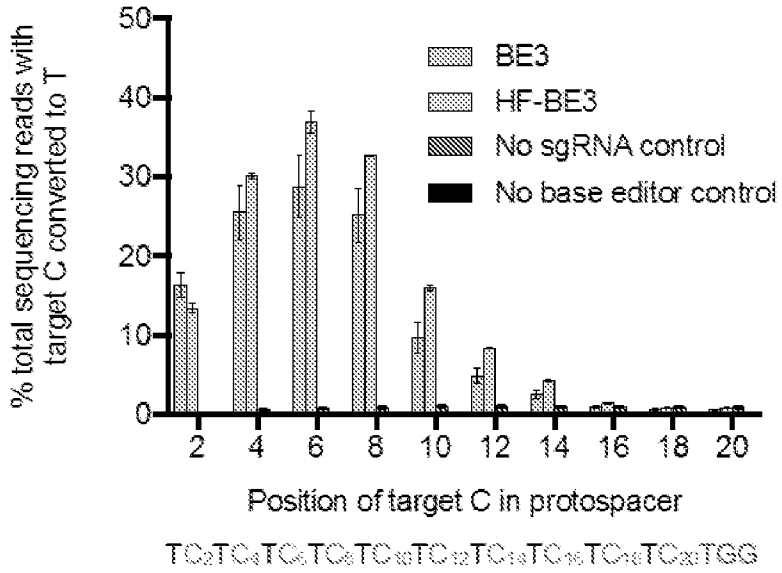


FIGURE 125B

Protein	Cas9	BE3	HF-BE3	BE3	HF-BE3	BE3	HF-BE3
E.Coli cell type	Rosetta	Rosetta	Rosetta	BL21 Star	BL21 Star	Rosetta pLysS	Rosetta pLysS

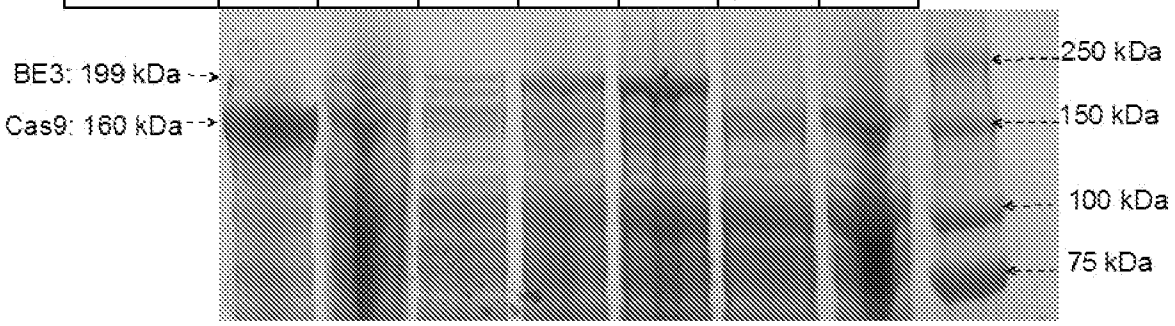


FIGURE 126A

Protein	HF-BE3	BE3	HF-BE3	BE3	HF-BE3	BE3	Cas9	Cas9
His-tag	N-term	N-term	C-term	C-term	C-term	C-term	N-term	N-term
Induction OD_{600}	0.7	0.7	0.4	0.7	0.7	0.4	0.4	0.4
Purification method	Nickel resin + FPLC	Nickel resin + FPLC	Nickel resin + FPLC	Nickel resin + FPLC	Nickel resin + FPLC	Nickel resin + FPLC	Nickel resin	Nickel resin + FPLC

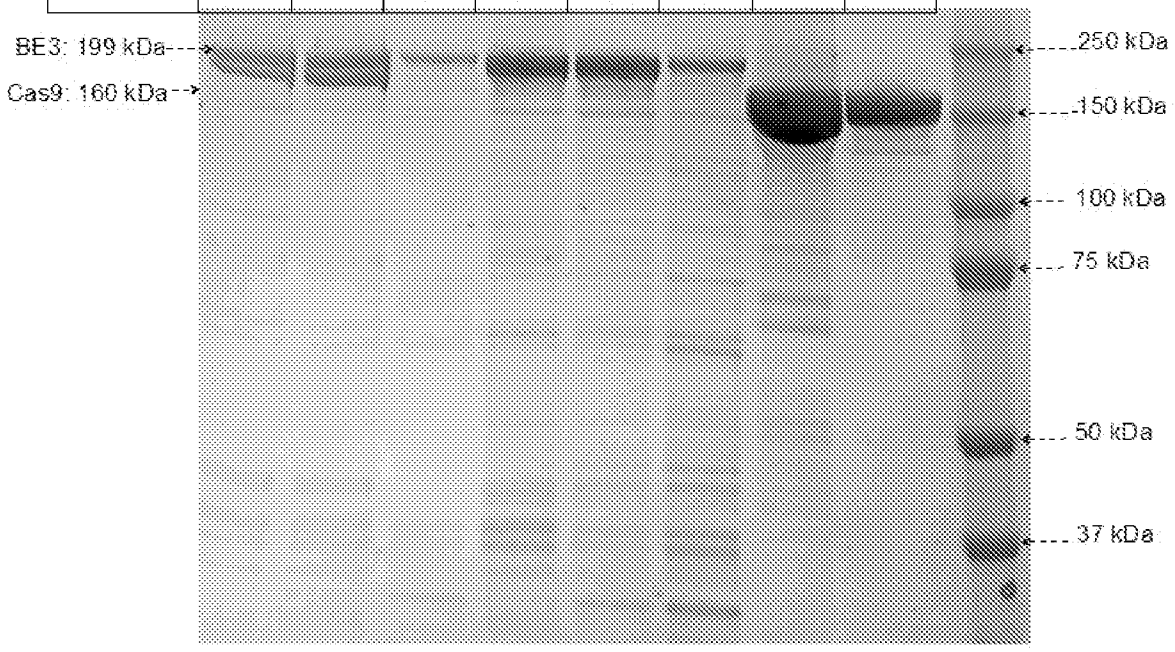


FIGURE 126B

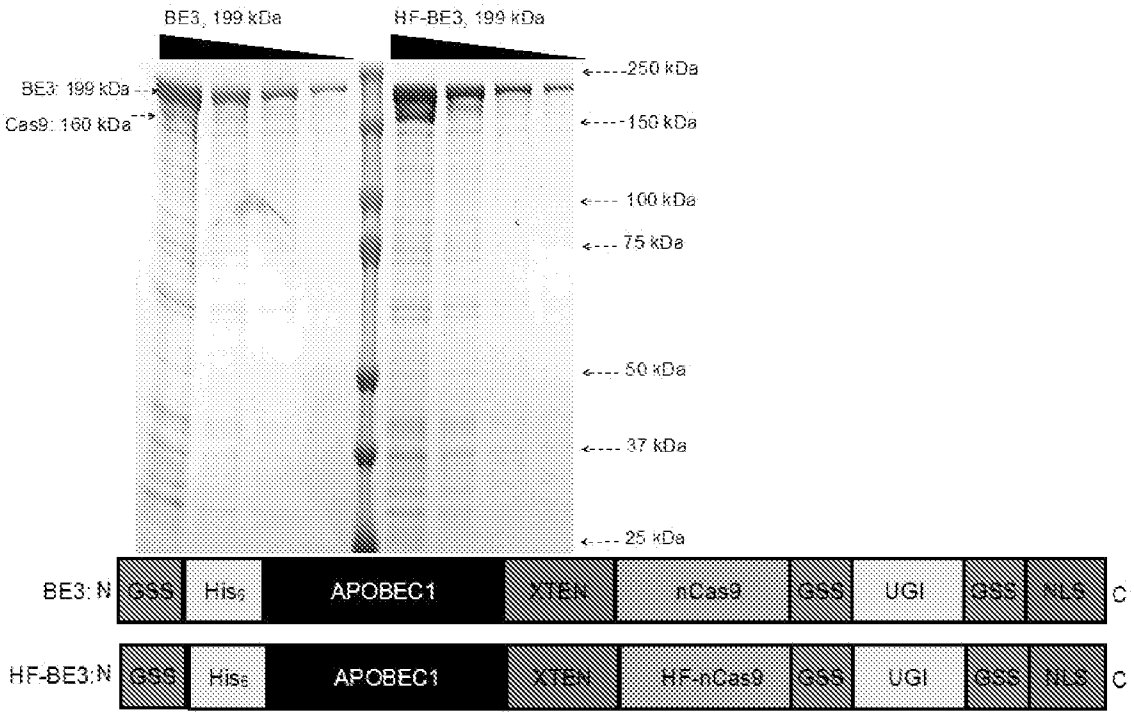


FIGURE 126C

EMX1 on-target : GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

EMX1 off-target 1: GAGTC$_5$TAAGCAGAAGAAGAAGAG

EMX1 off-target 2: GAGGC$_5$C$_6$GAGCAGAAGAAAGACGG

EMX1 off-target 3: GAGTC$_5$C$_6$TAGCAGGAGAAGAAGAG

FIGURE 127A

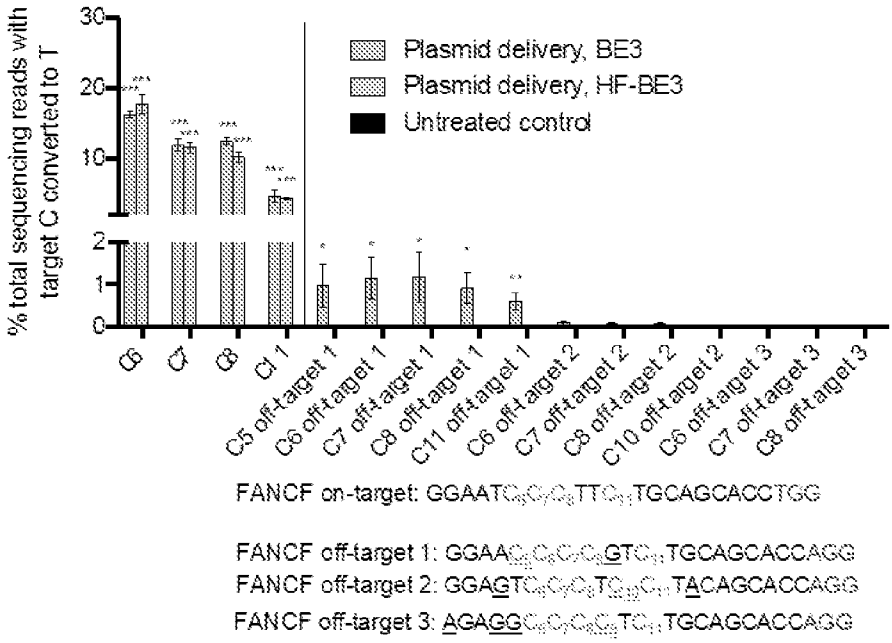


FIGURE 127B

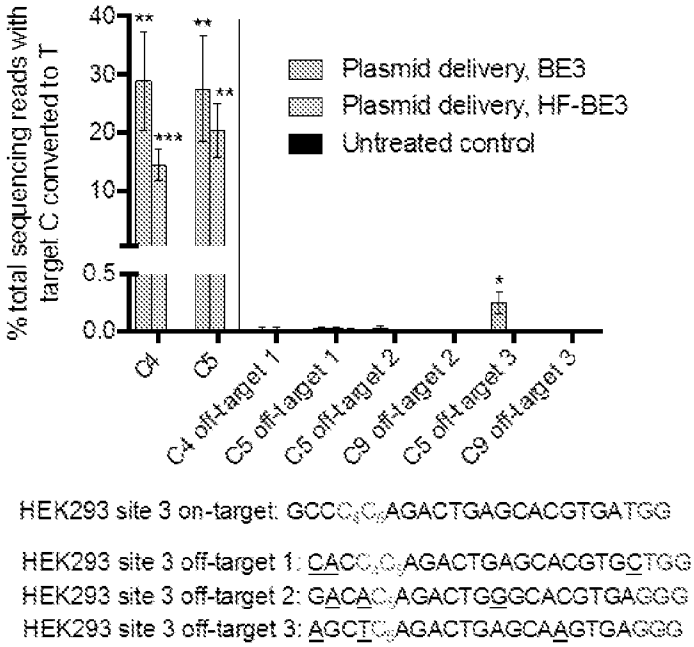


FIGURE 127C

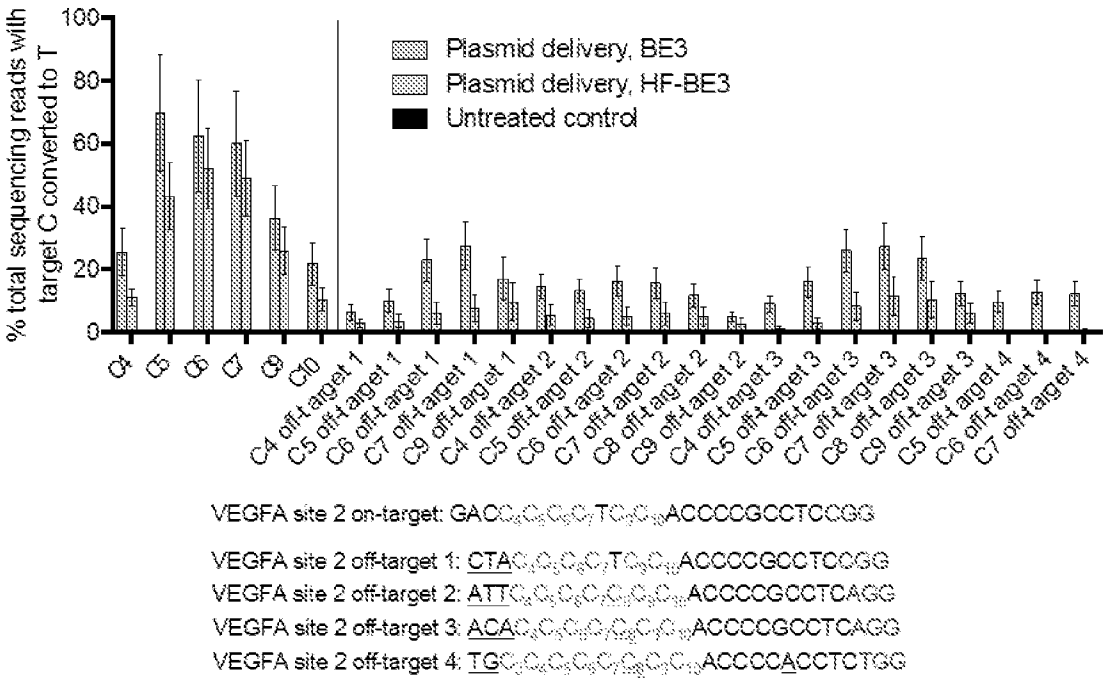


FIGURE 127D

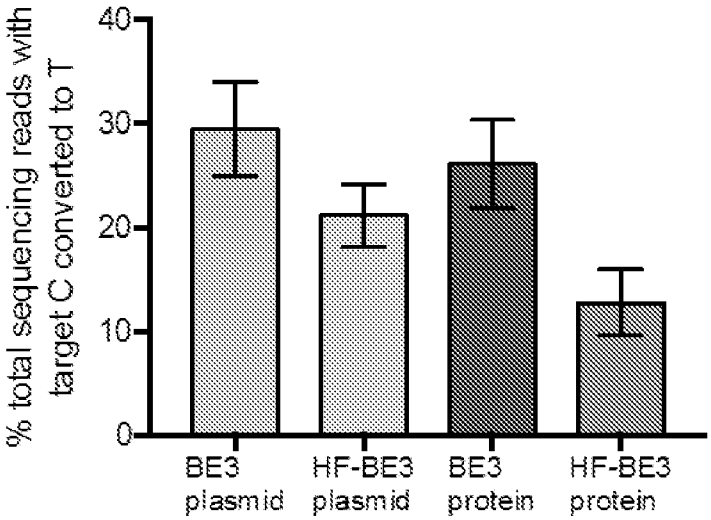
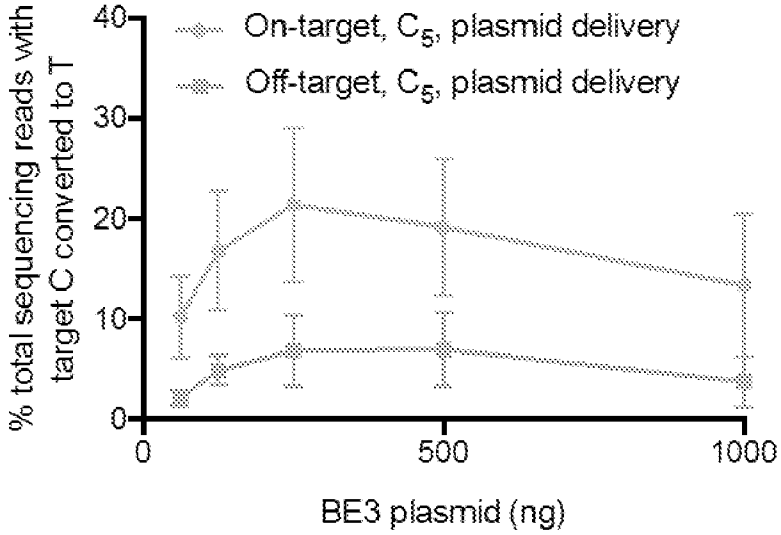


FIGURE 128A



EMX1: GAGTC5C6GAGCAGAAGAAGAAGGG
EMX1 off target 1: GAGTC5TAAGCAGAAGAAGAAGAG

FIGURE 128B

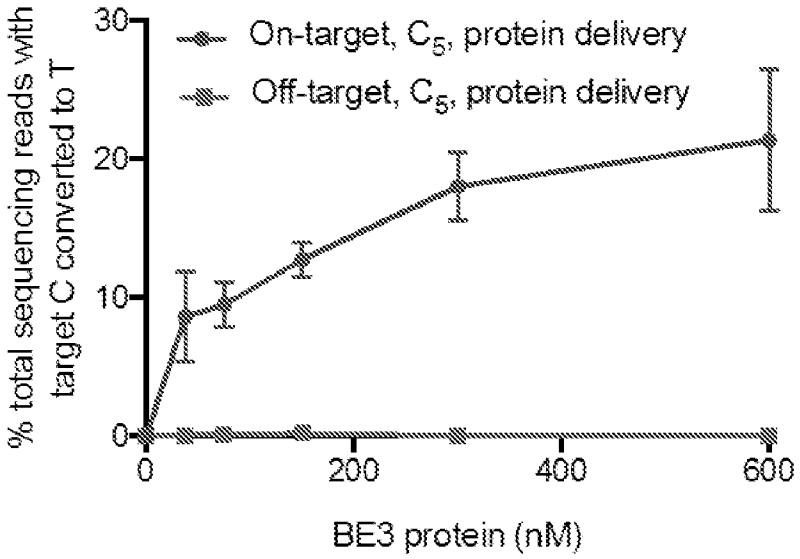


FIGURE 128C

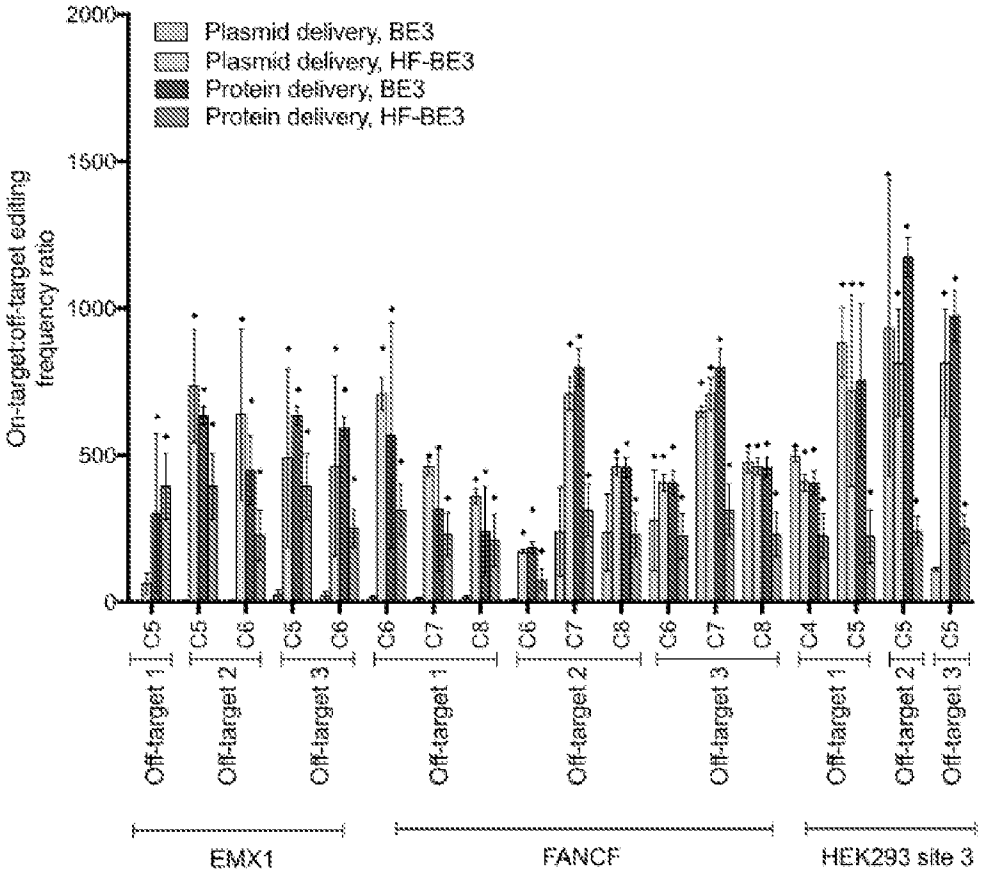


FIGURE 129A

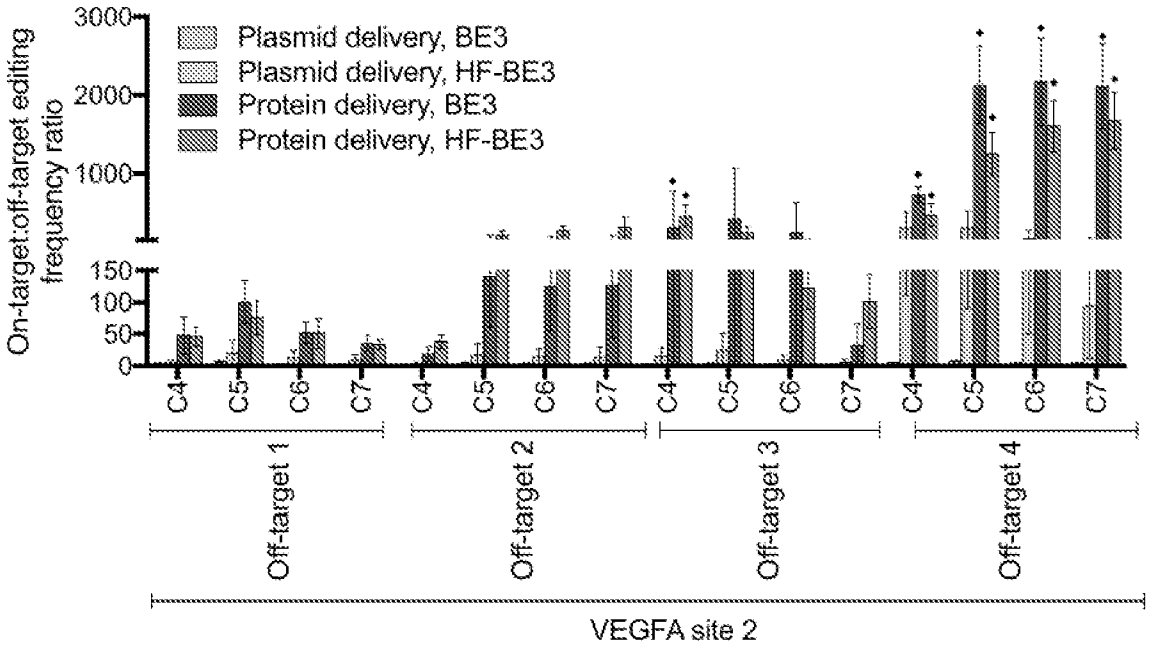
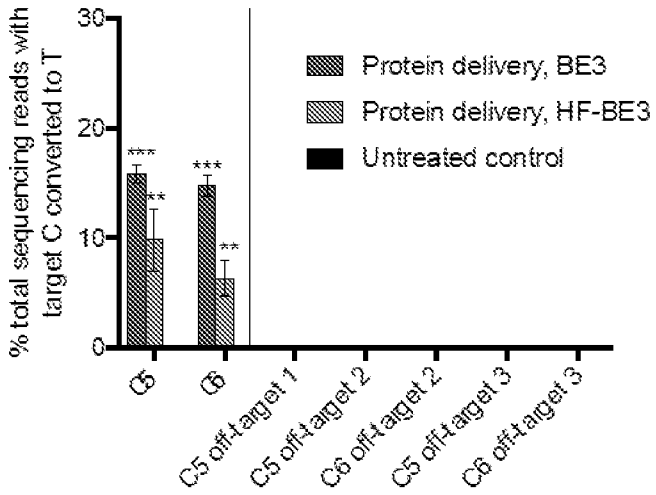


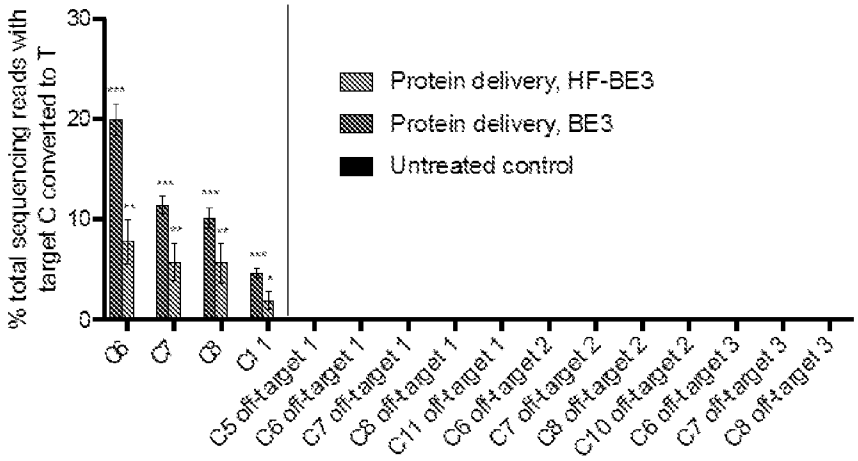
FIGURE 129B



EMX1 on-target : GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

EMX1 off-target 1: GAGTC$_5$TAAGCAGAAGAAGAAGAG
EMX1 off-target 2: GAGGC$_5$C$_6$GAGCAGAAGAAAGACGG
EMX1 off-target 3: GAGTC$_5$C$_6$TAGCAGGAGAAGAAGAG

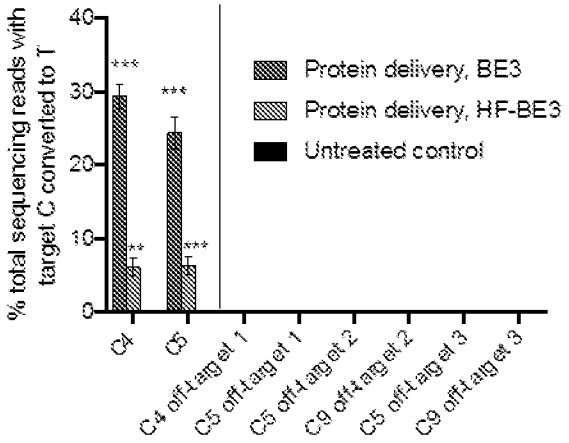
FIGURE 130A



FANCF on-target: GGAATC$_5$C$_6$C$_7$TTC$_{11}$TGCAGCACCTGG

FANCF off-target 1: GGAAC$_5$C$_6$C$_7$C$_8$GTC$_{11}$TGCAGCACCAGG
FANCF off-target 2: GGAGTC$_5$C$_6$C$_7$TC$_{10}$C$_{11}$TACAGCACCAGG
FANCF off-target 3: AGAGGC$_5$C$_6$C$_7$C$_8$TC$_{11}$TGCAGCACCAGG

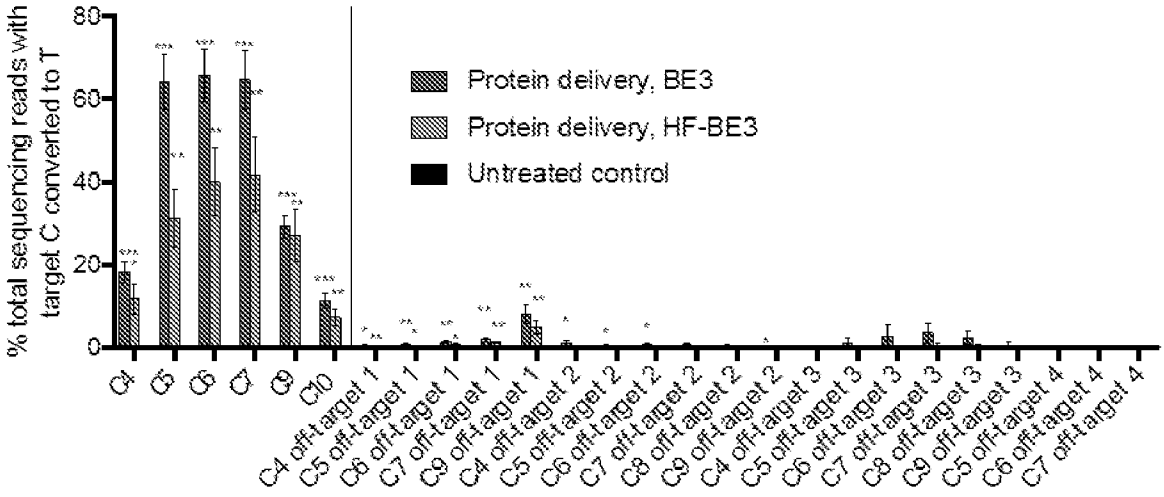
FIGURE 130B



HEK293 site 3 on-target: GCCC$_3$C$_5$AGACTGAGCACGTGATGG

HEK293 site 3 off-target 1: CAC$_1$C$_3$C$_5$AGACTGAGCACGTGCTGG
HEK293 site 3 off-target 2: GACAC$_5$AGACTGGGCACGTGAGGG
HEK293 site 3 off-target 3: AGCTC$_5$AGACTGAGCAAGTGAGGG

FIGURE 130C



VEGFA site 2 on-target: GACC$_3$C$_4$C$_5$C$_6$C$_7$TC$_9$C$_{10}$ACCCCGCCTCCGG

VEGFA site 2 off-target 1: CTAC$_3$C$_4$C$_5$C$_6$C$_7$TC$_9$C$_{10}$ACCCCGCCTCCGG
VEGFA site 2 off-target 2: ATTC$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCCCGCCTCAGG
VEGFA site 2 off-target 3: ACAC$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCCCGCCTCAGG
VEGFA site 2 off-target 4: TGC$_3$C$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCCCACCTCTGG

FIGURE 130D

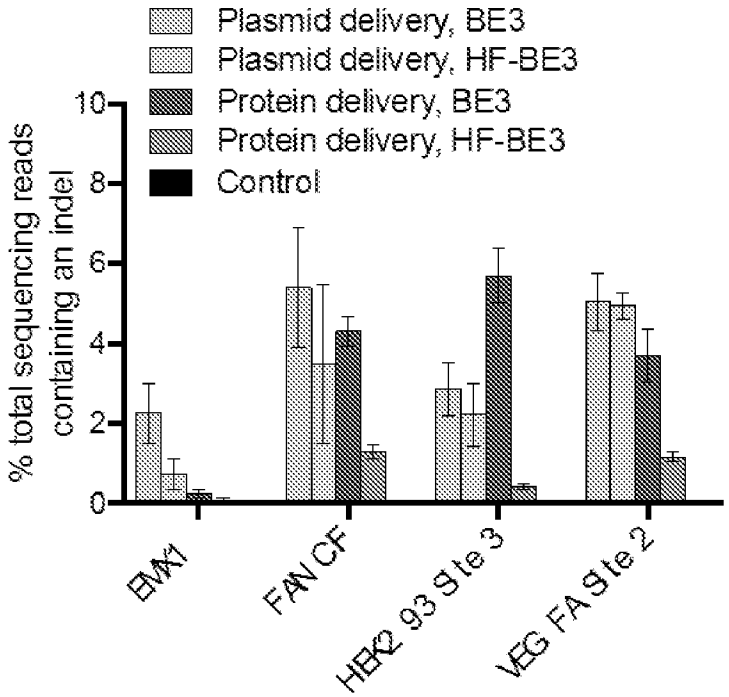


FIGURE 131A

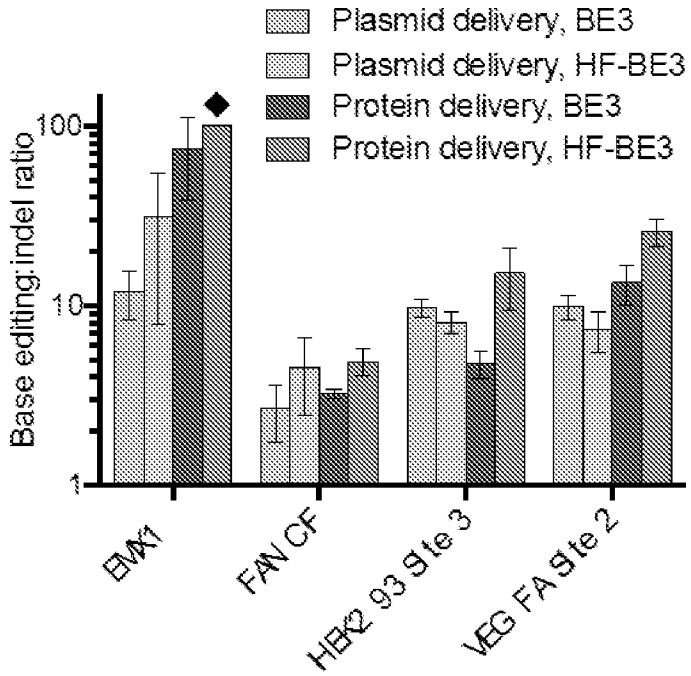


FIGURE 131B

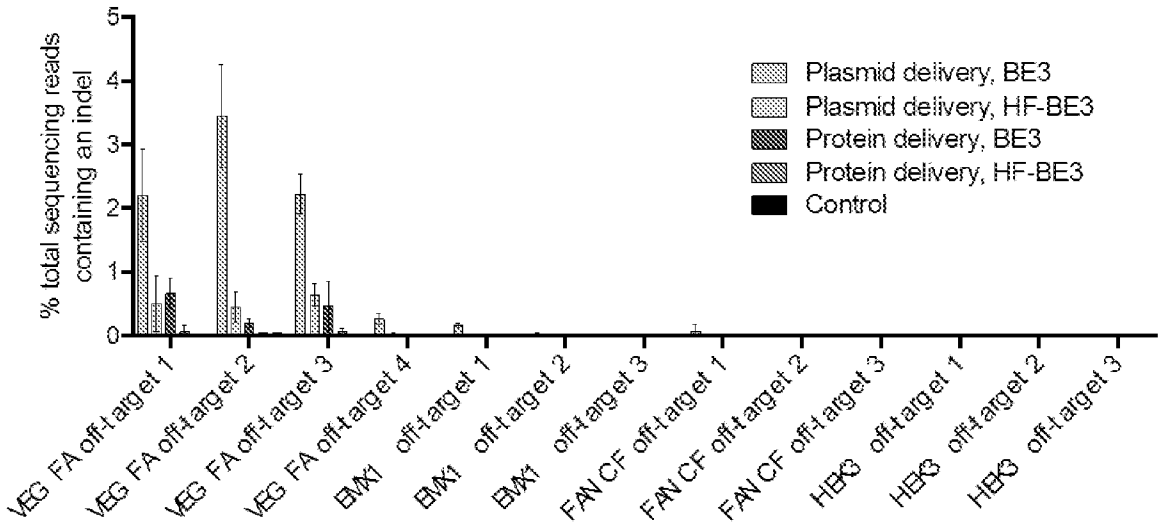
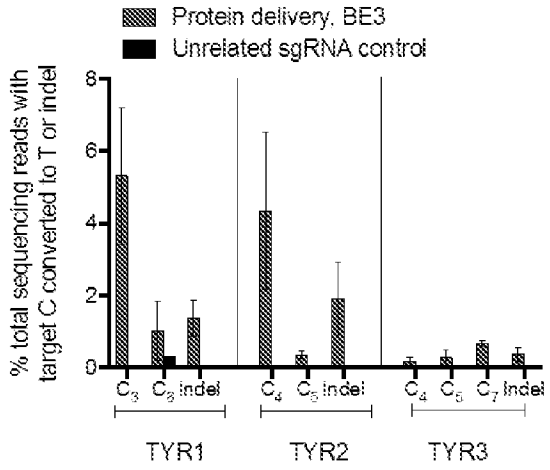


FIGURE 131C



TYR1: GTC$_3$AGGTC$_9$GAGGGTTCTGTCAGG
TYR2: CTTC$_4$C$_5$AGGATGAGAACACAGAGG
TYR3: CAAC$_4$C$_5$AC$_7$TGCTCAAAGATGCTGG

FIGURE 132A

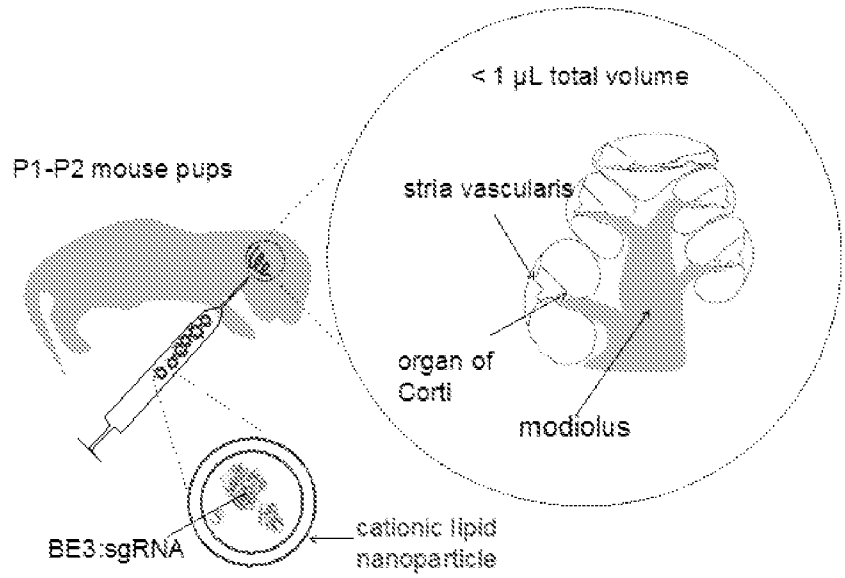
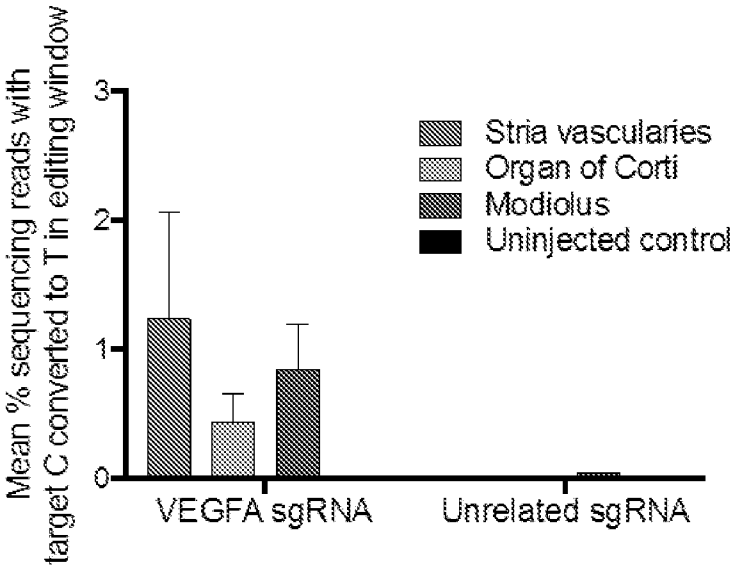


FIGURE 132B



VEGFA site 2: GACC$_4$C$_5$C$_6$C$_7$TC$_9$C$_{10}$ACCCCGCCTCCGG

FIGURE 132C

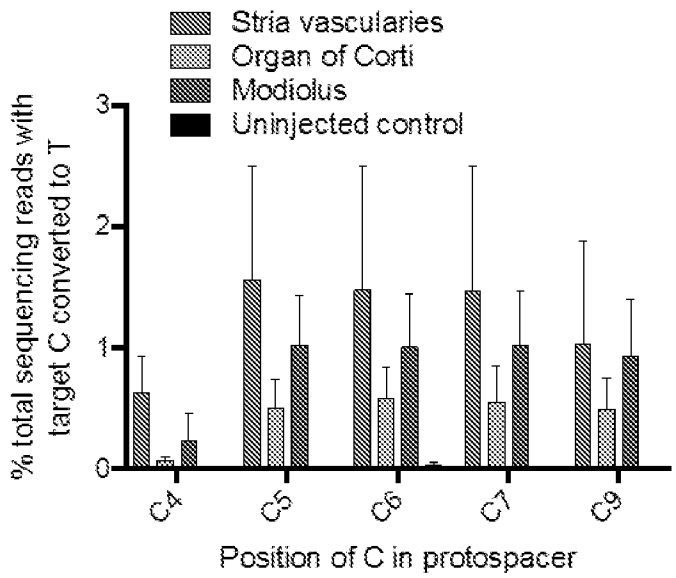


FIGURE 132D

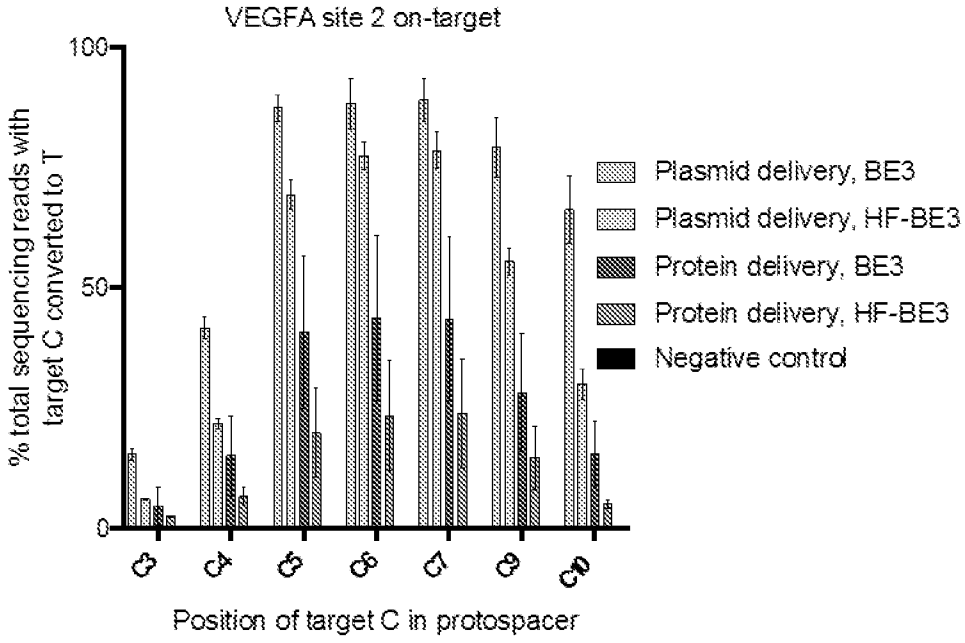


FIGURE 133A

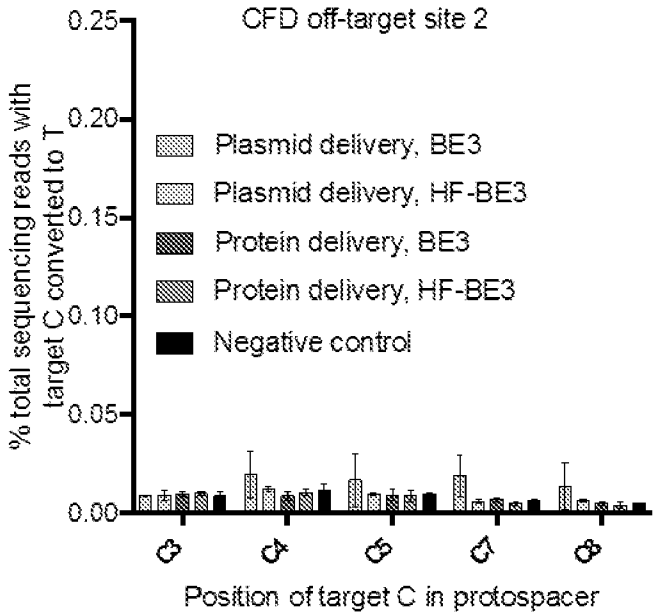


FIGURE 133B

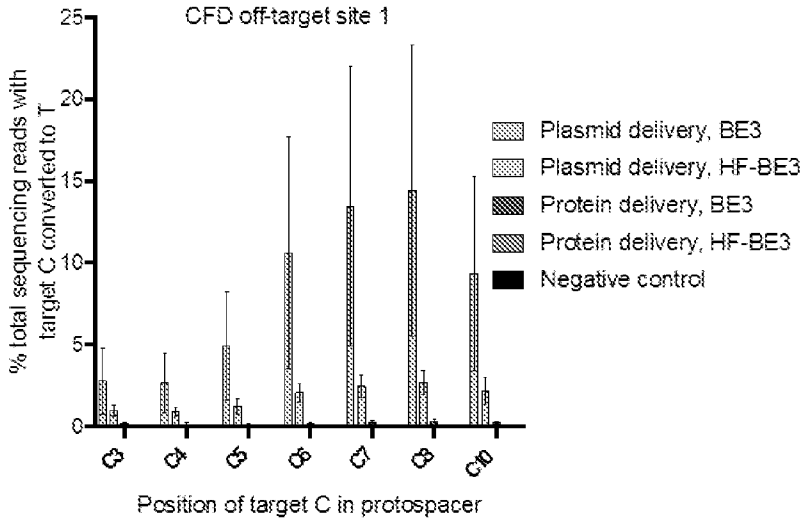


FIGURE 133C

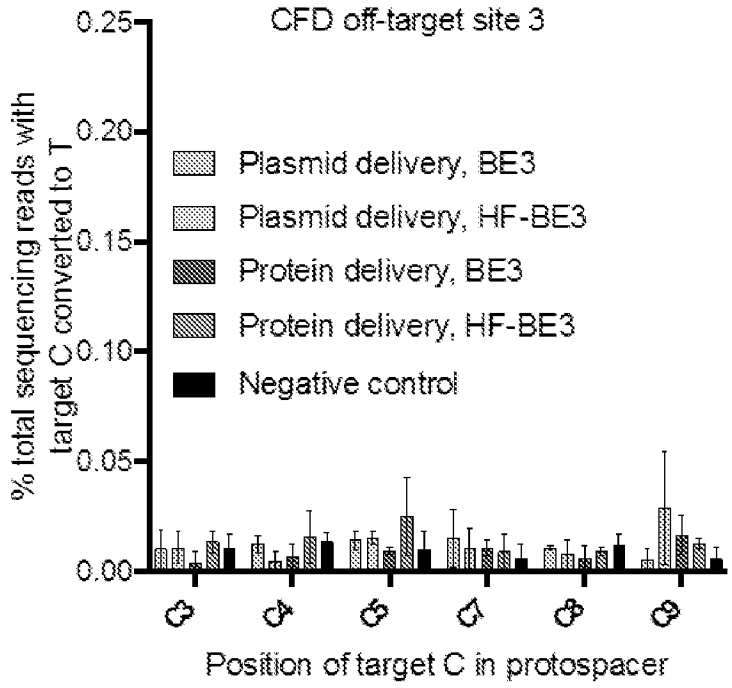


FIGURE 133D

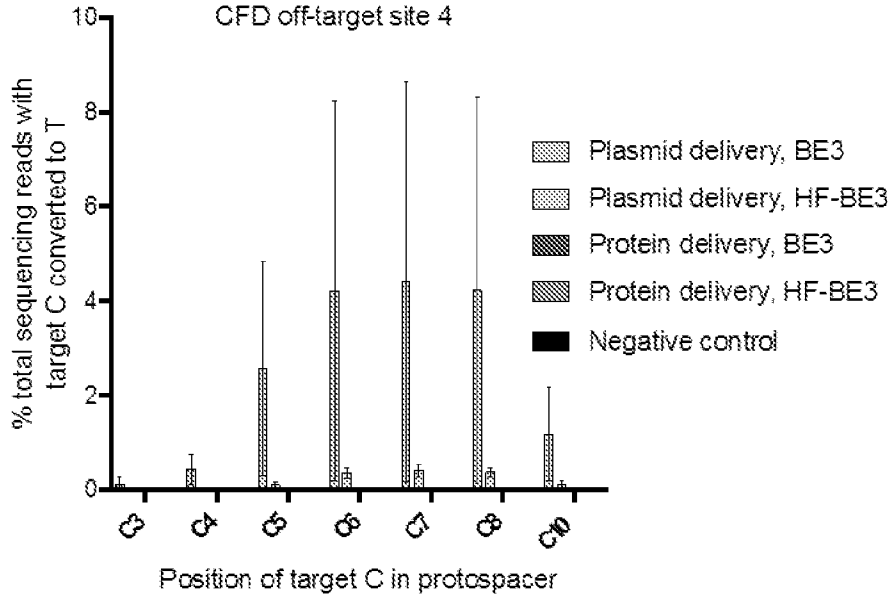


FIGURE 133E

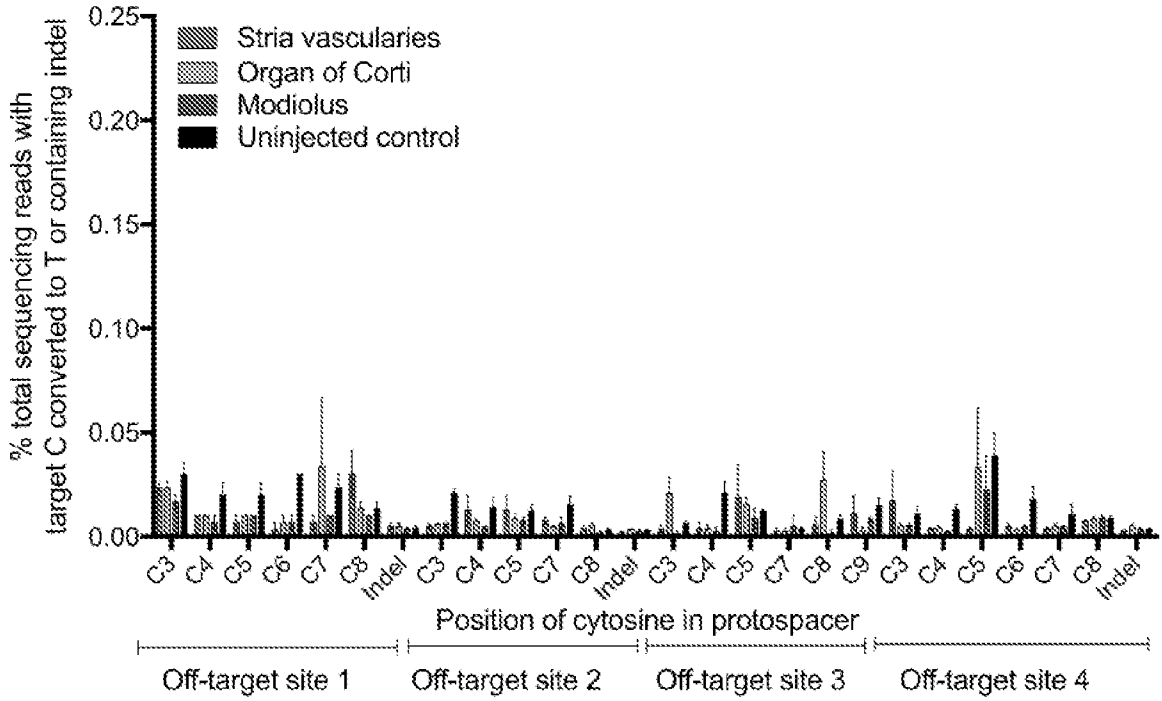


FIGURE 134A

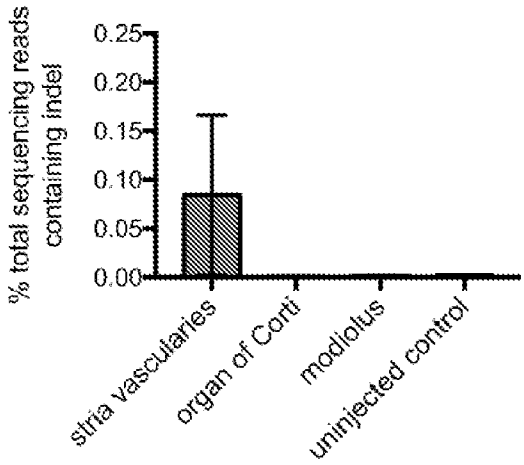


FIGURE 134B

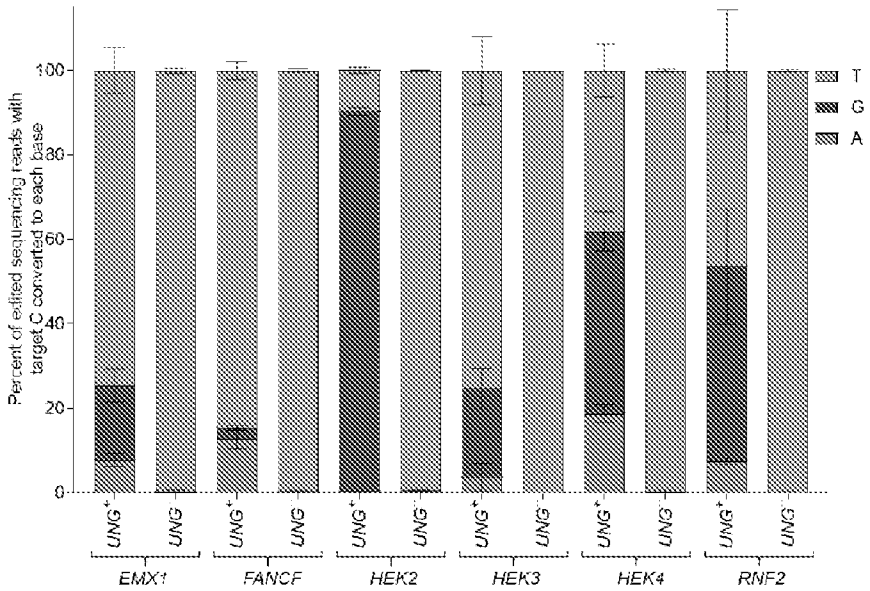


FIGURE 135A

EMX1: GAGTC$_5$CGAGCAGAAGAAGAAGGG

FANCF: GGAATCC$_7$CTTCTGCAGCACCTGG

HEK2: GAACAC$_5$AAAGCATAGACTGCGGG

HEK3: GGCCC$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTCATC$_6$TTAGTCATTACCTGAGG

FIGURE 135B

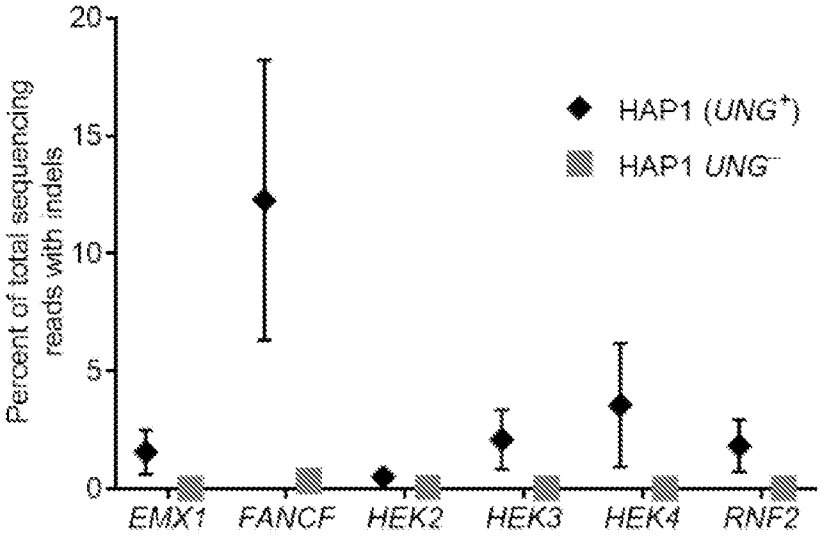


FIGURE 135C

HEK4		G	C	C	A	C	T	G	C	G	C	T	G	G	A	G	G	T	G	G	G	G
untreated	A	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.4	99.9	99.9	100.0	99.9	99.9
	T	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.5	0.1	0.0	0.0	0.0	0.0

BE3		A	C	G	T	A	C	T	G	G	C	T	G	G	A	G	G	T	G	G	G	G
	A	0.1	0.0	0.1	99.9	8.4	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	99.0	0.0	41.9	0.0	0.0	99.3	0.0	0.0	99.8	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	24.5	0.0	100.0	0.0	100.0	99.9	0.0	0.0	100.0	100.0	0.1	100.0	100.0	0.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	0.9	0.0	25.2	99.9	0.0	0.5	0.0	0.2	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0

AID-BE3		A	C	G	T	A	C	T	G	G	C	T	G	G	A	G	G	T	G	G	G	G
	A	0.3	0.4	0.3	100.0	0.9	0.0	0.0	0.9	0.0	0.0	0.1	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	61.5	0.0	71.3	0.0	0.0	80.7	0.0	0.0	97.9	0.1	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0
	G	99.7	99.6	0.1	0.0	0.7	0.0	100.0	0.1	100.0	100.0	0.0	0.0	99.9	100.0	0.0	100.0	100.0	0.2	99.9	100.0	100.0	99.9
	T	0.0	0.0	18.0	0.0	27.1	100.0	0.0	18.3	0.0	0.0	2.0	99.9	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.1

FIGURE 136A

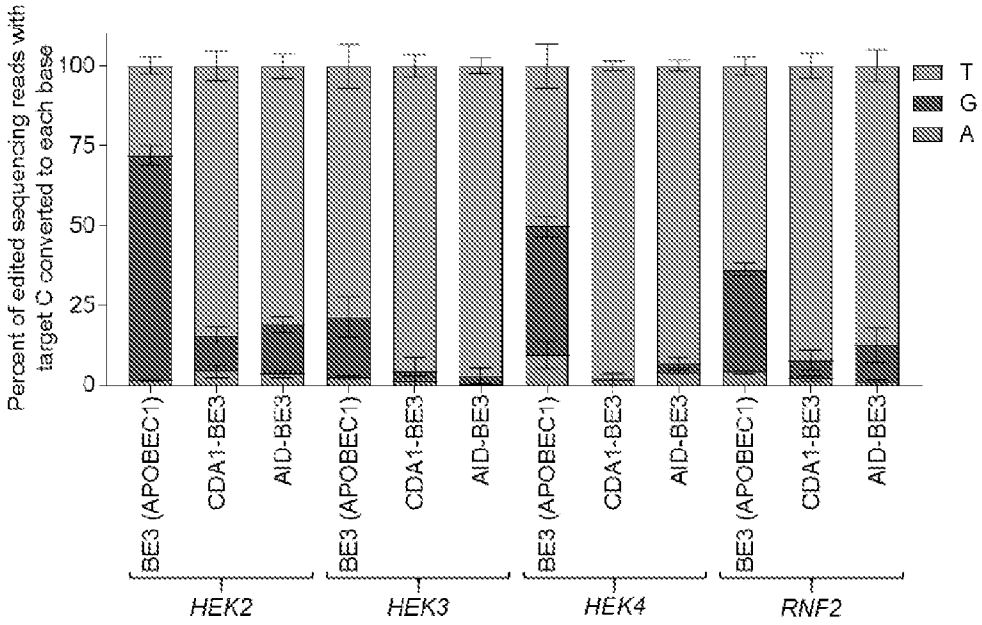


FIGURE 136B

HEK2: GAACAC$_9$AAAGCATAGACTGCGGG

HEK3: GGCCC$_9$AGACTGAGCACGTGATGG

HEK4: GGCAC$_9$TGCGGCTGGAGGTCCGGG

RNF2: GTCATC$_9$TTAGTCATTACCTGAGG

FIGURE 136C

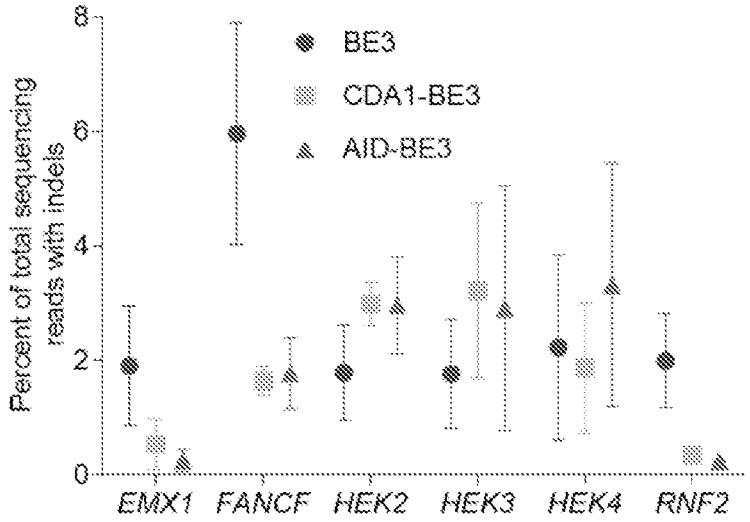


FIGURE 136D

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

FIGURE 137A

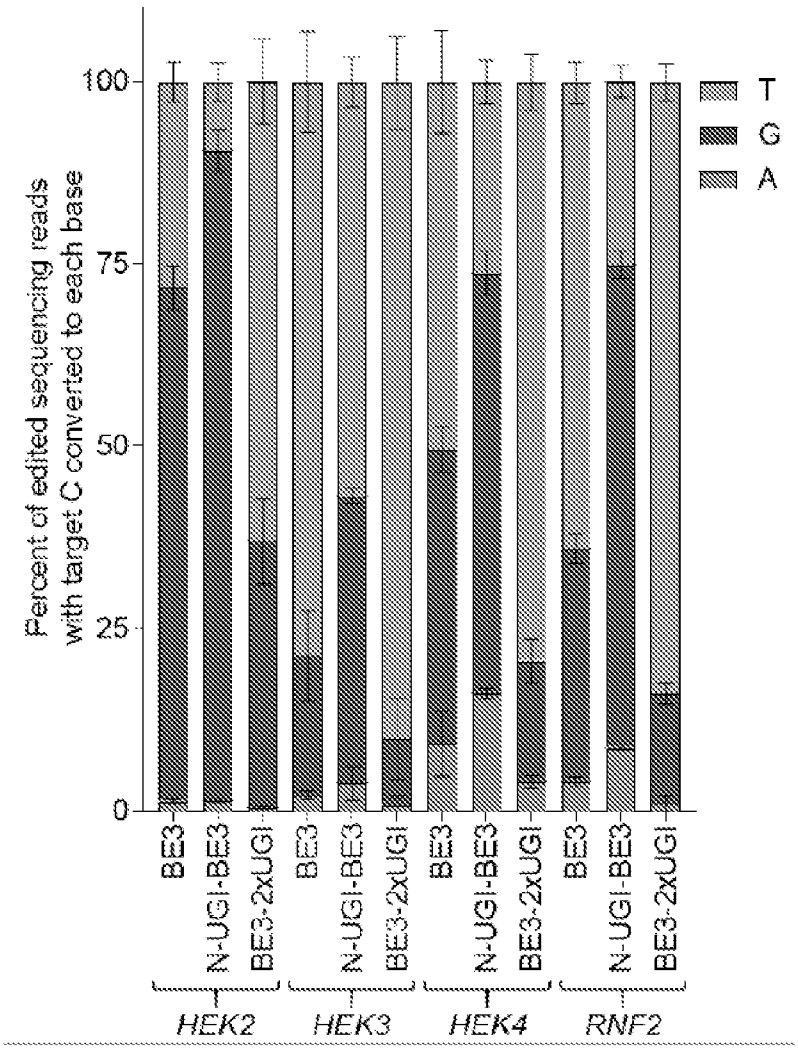


FIGURE 137B

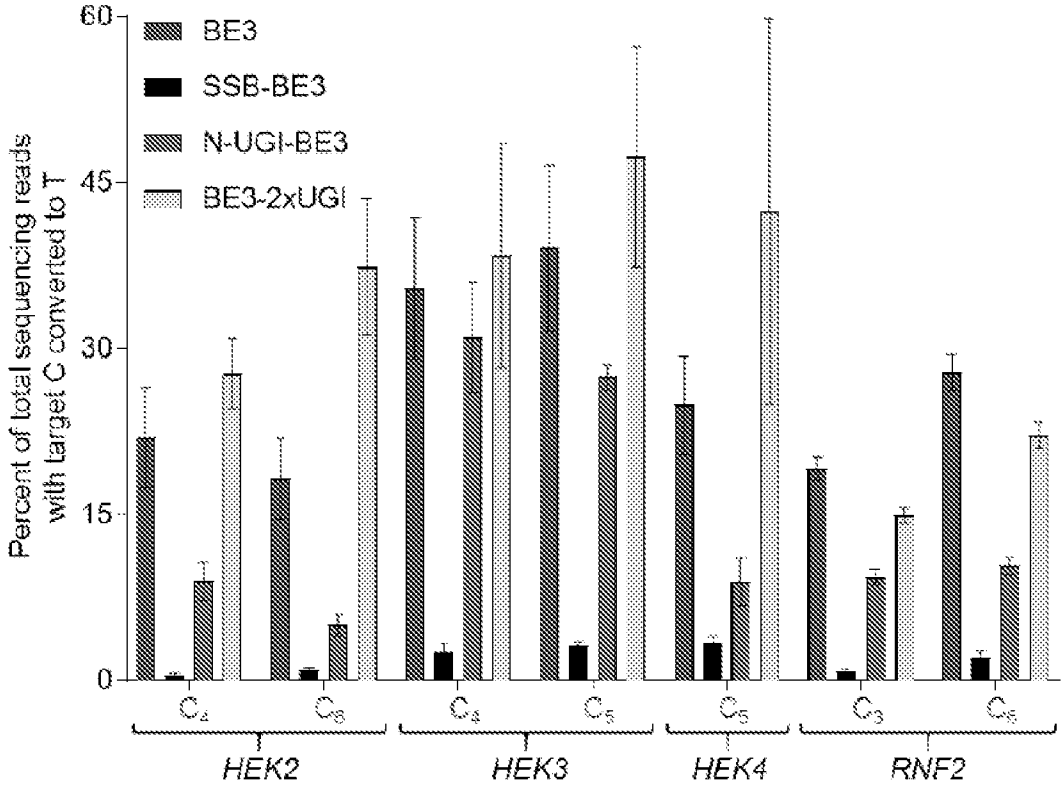

60
45
30
15
0
BE3
SSB-BE3
N-UGI-BE3
BE3-2xUGI
Percent of total sequencing reads with target C converted to T
C4
C8
C4
C5
C5
C3
C5
HEK2
HEK3
HEK4
RNF2

FIGURE 137C

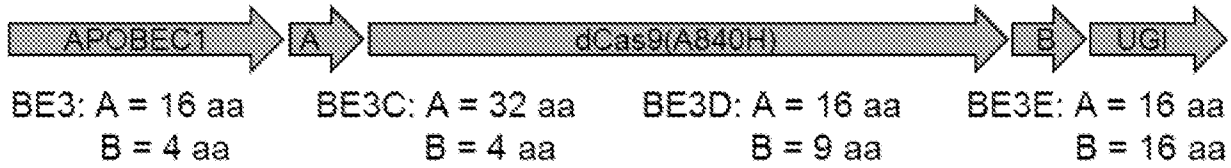


FIGURE 138A

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_8$TTAGTCATTACCTGAGG

FIGURE 138B

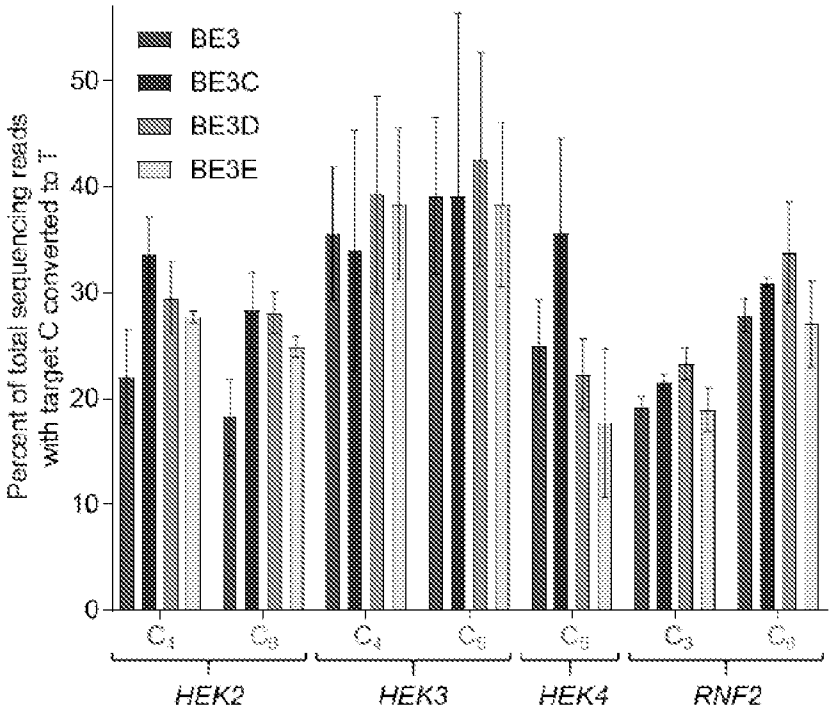


FIGURE 138C

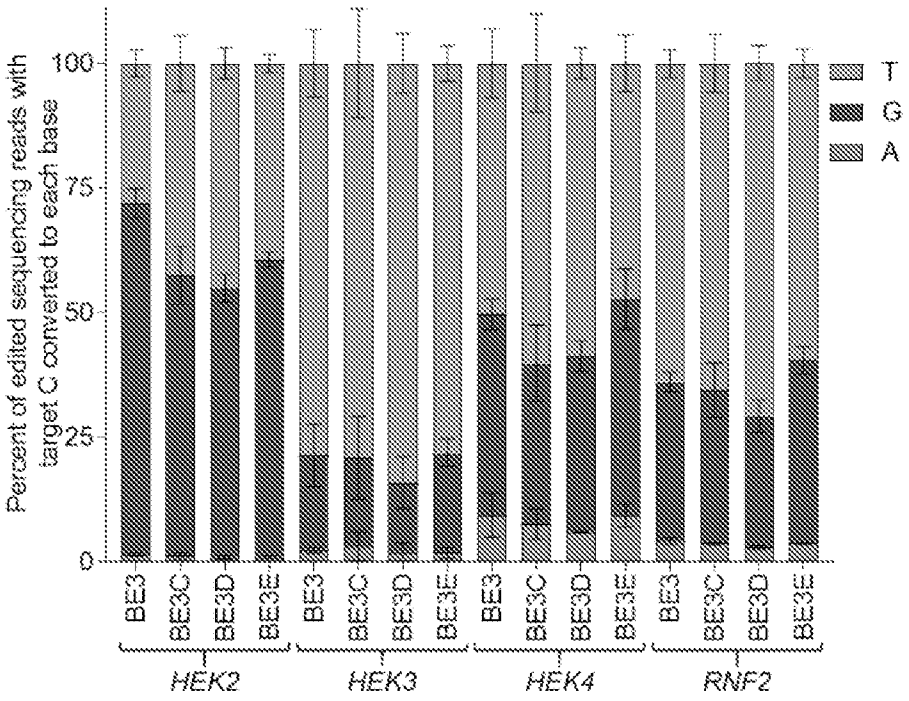

Percent of edited sequencing reads with target C converted to each base
100
75
50
25
0
T
G
A
BE3
BE3C
BE3D
BE3E
HEK2
HEK3
HEK4
RNF2

FIGURE 138D

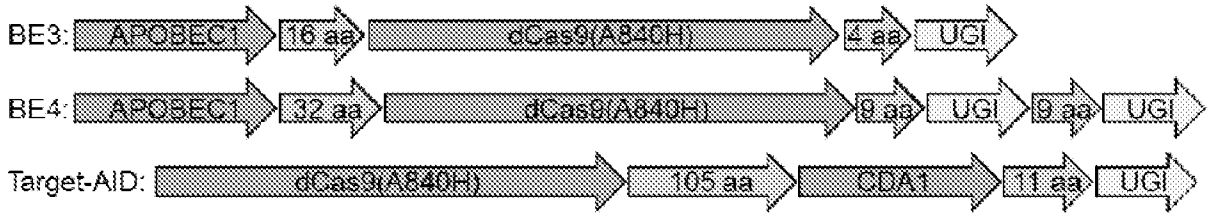


FIGURE 139A

EMX1:	GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG
FANCF:	GGAATC$_6$C$_7$C$_8$TTCTGCAGCACCTGG
HEK2:	GAAC$_4$AC$_6$AAAGCATAGACTGCGGG
HEK3:	GGCC$_4$C$_5$AGACTGAGCACGTGATGG
HEK4:	GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2:	GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

FIGURE 139B

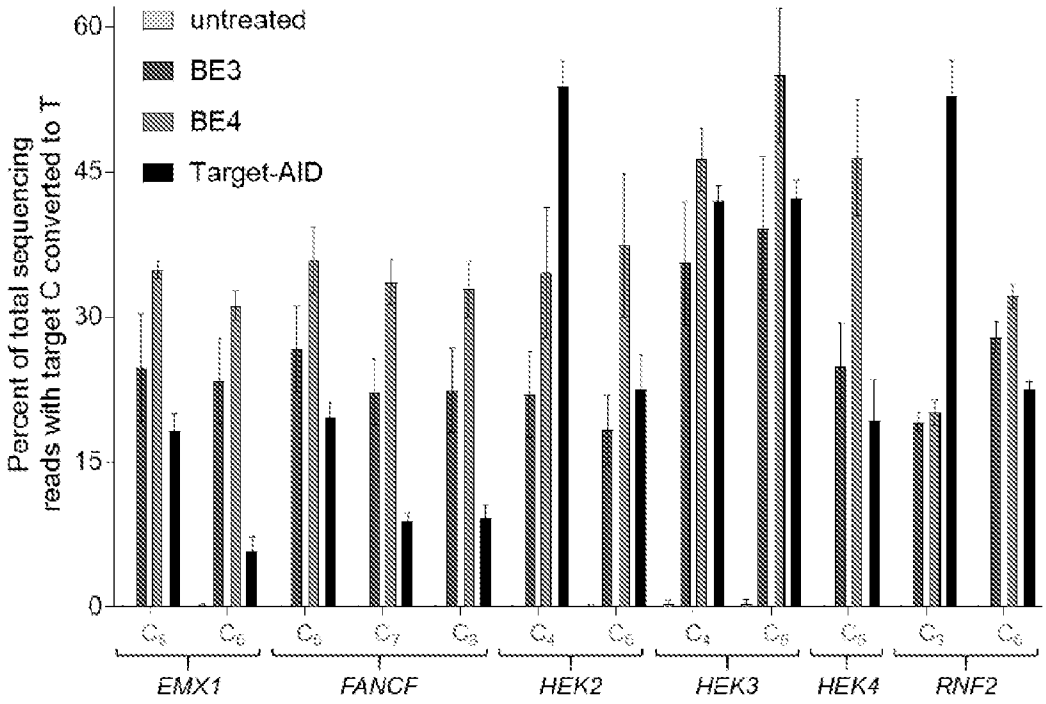


FIGURE 139C

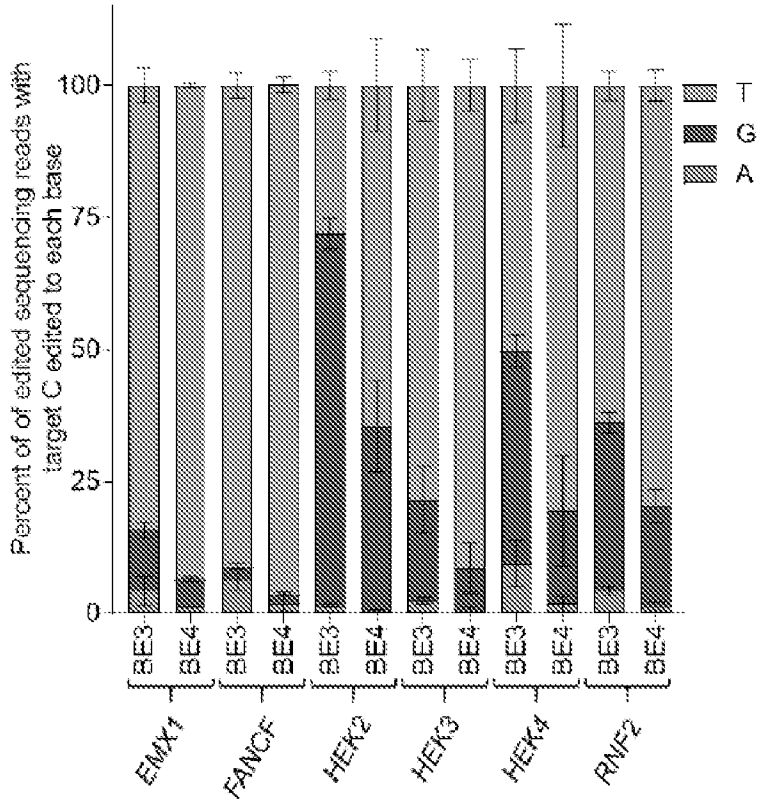

Percent of of edited sequencing reads with target C edited to each base
100
75
50
25
0
T
G
A
BE3
BE4
EMX1
FANCF
HEK2
HEK3
HEK4
RNF2

FIGURE 139D

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

FANCF: GGAATC$_6$C$_7$C$_8$TTCTGCAGCACCTGG

HEK2: GAAC$_5$AC$_8$AAAGCATAGACTGCGGG

HEK3: GGC$_3$C$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGC$_3$AC$_5$TGC$_8$GGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

FIGURE 140A

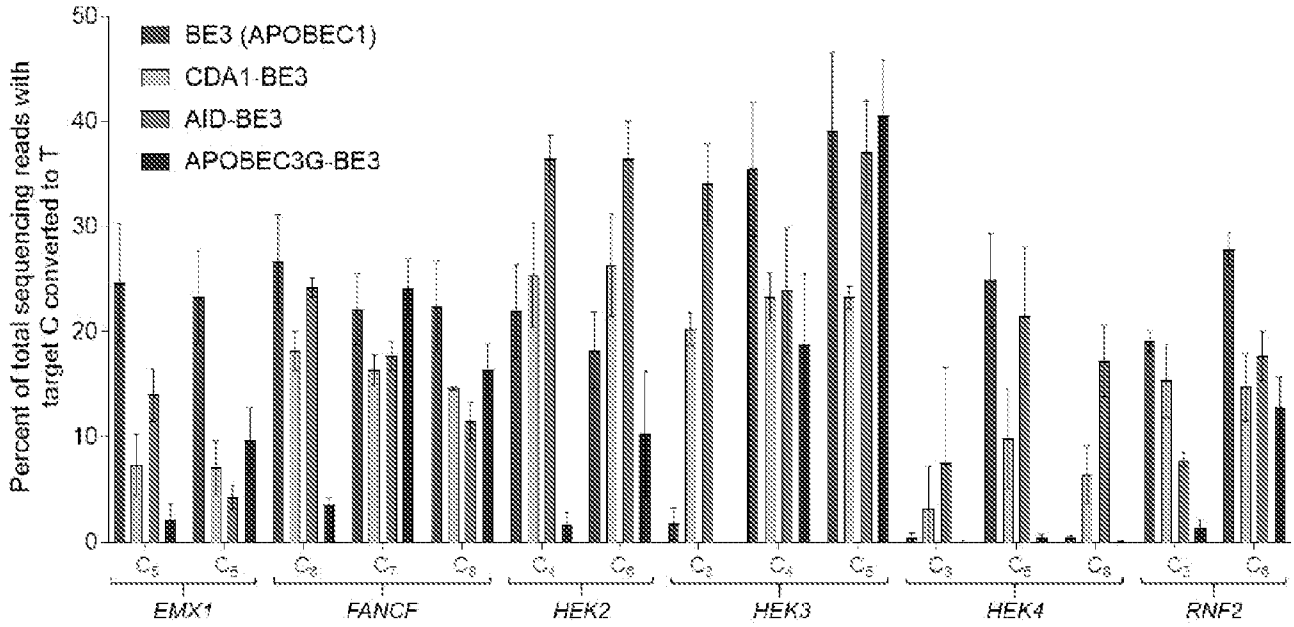


FIGURE 140B

	Percent of cleanly edited reads with given sequence		
HEK2 protospacer sequence	BE3	CDA1-BE3	AID-BE3
GAAT$_5$AT$_8$AAAGCATAGACTGCGGG	77 ± 1	88 ± 2	92 ± 2
GAAC$_5$AT$_8$AAAGCATAGACTGCGGG	23 ± 1	5 ± 1	4 ± 1
GAAT$_5$AC$_8$AAAGCATAGACTGCGGG	0 ± 0	7 ± 2	4 ± 2

FIGURE 140C

EMX1: GAGTC$_3$C$_6$GAGCAGAAGAAGAAGGG

FANCF: GGAATC$_6$C$_7$C$_9$TTCTGCAGCACCTGG

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

FIGURE 141A

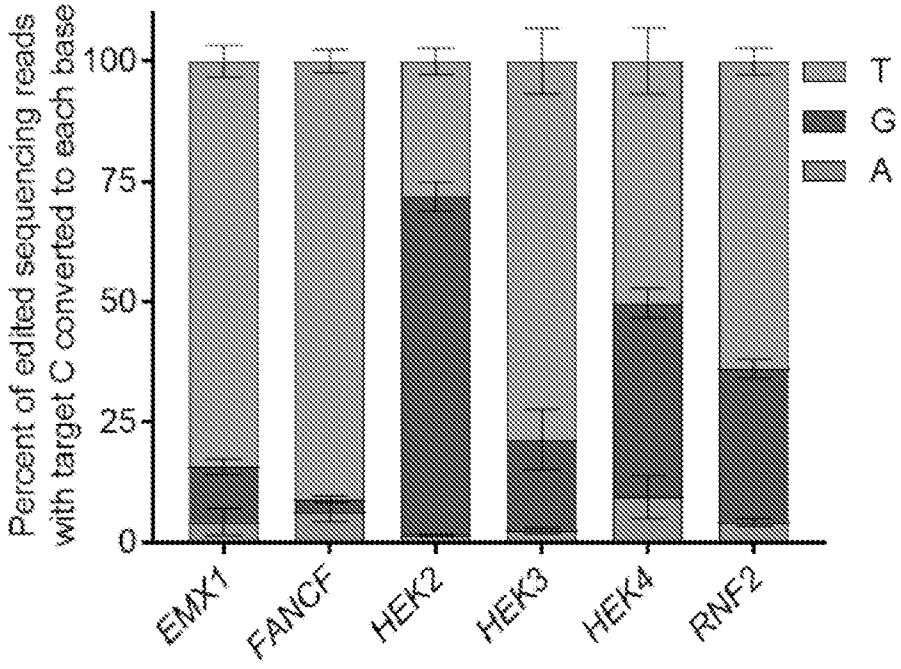


FIGURE 141B

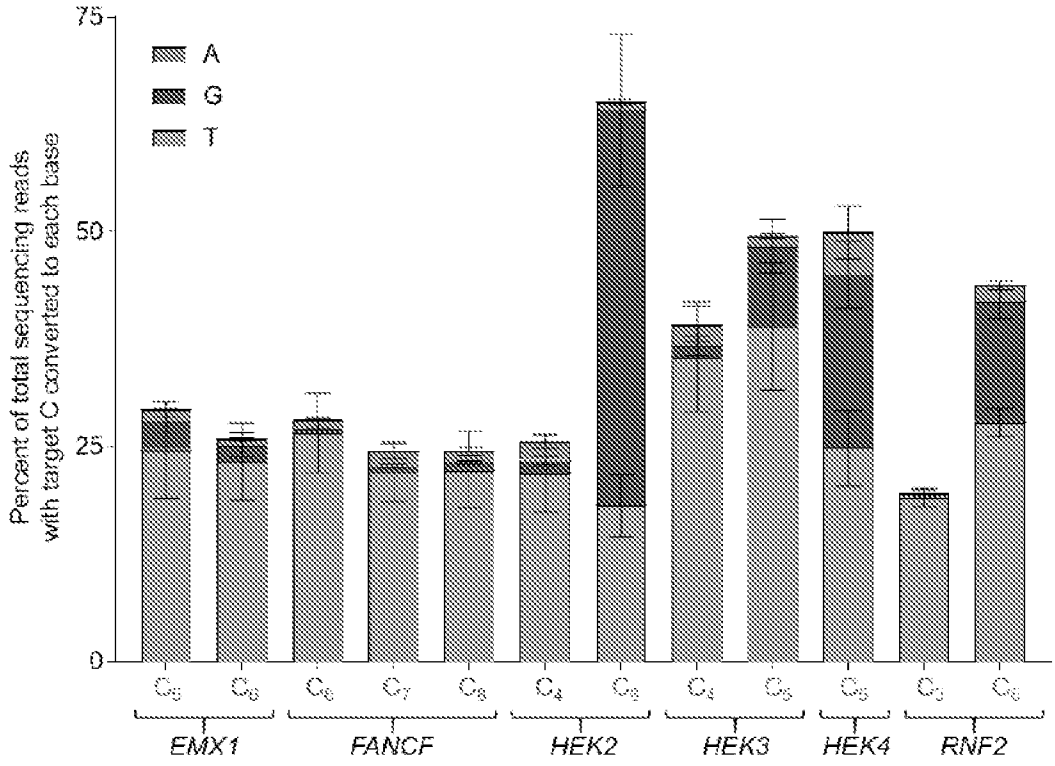

75
50
25
0
A
G
T
Percent of total sequencing reads with target C converted to each base
C5
C6
C6
C7
C8
C4
C5
C4
C5
C5
C3
C6
EMX1
FANCF
HEK2
HEK3
HEK4
RNF2

FIGURE 141C

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_8$TTAGTCATTACCTGAGG

FIGURE 142A

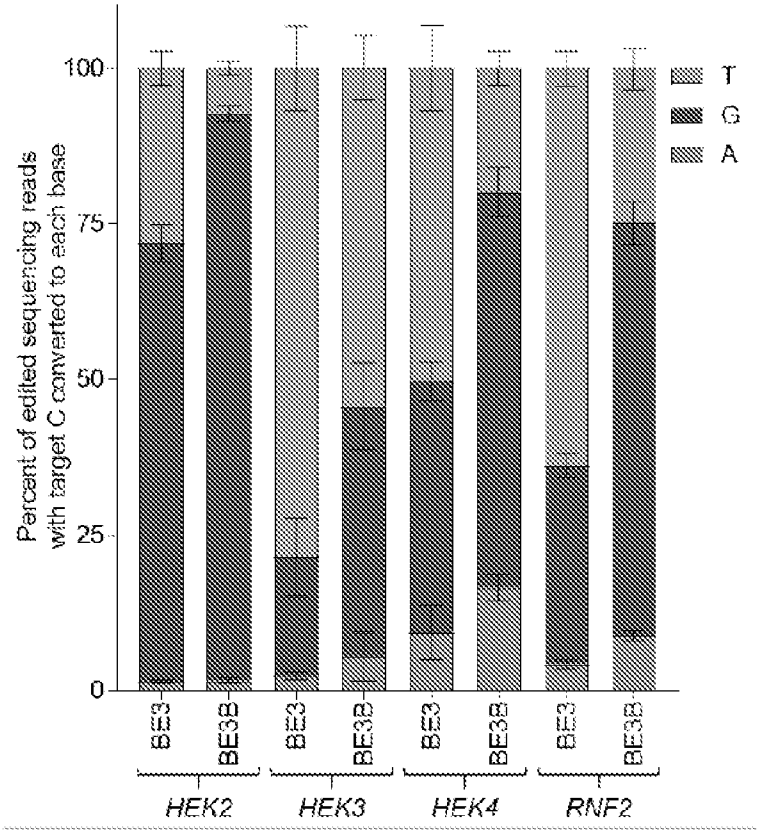


FIGURE 142B

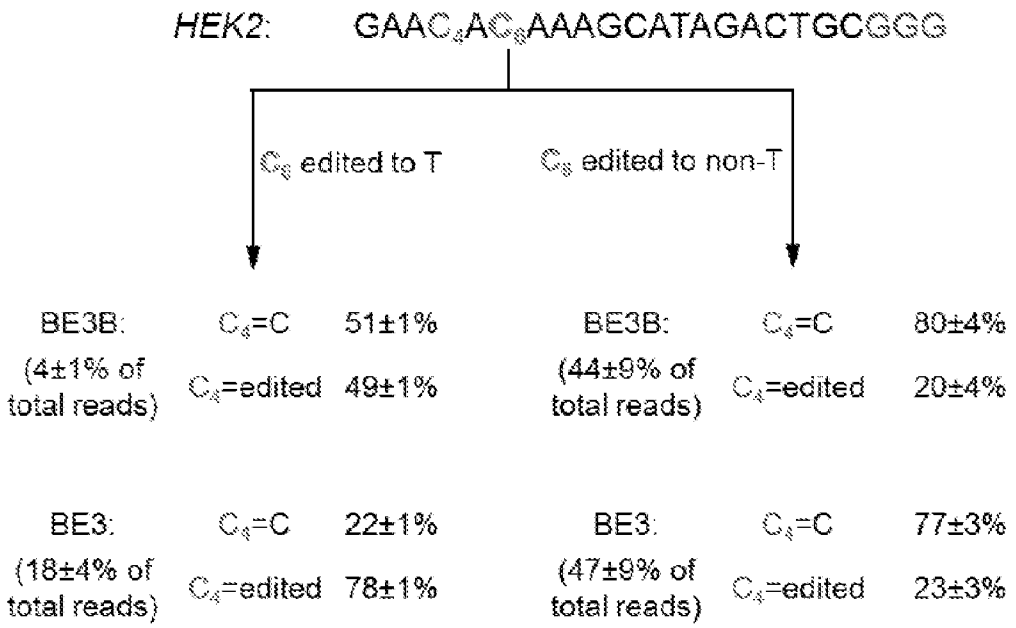


FIGURE 142C

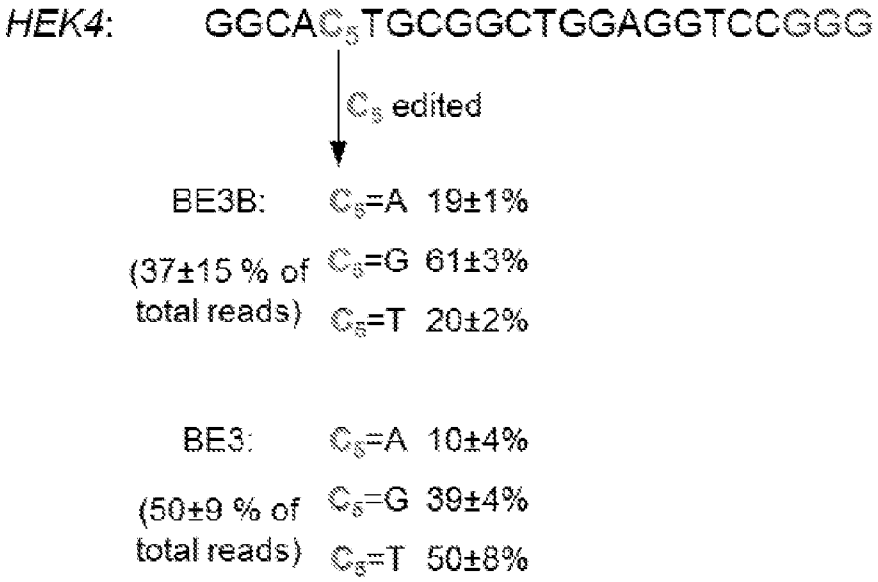


FIGURE 142D

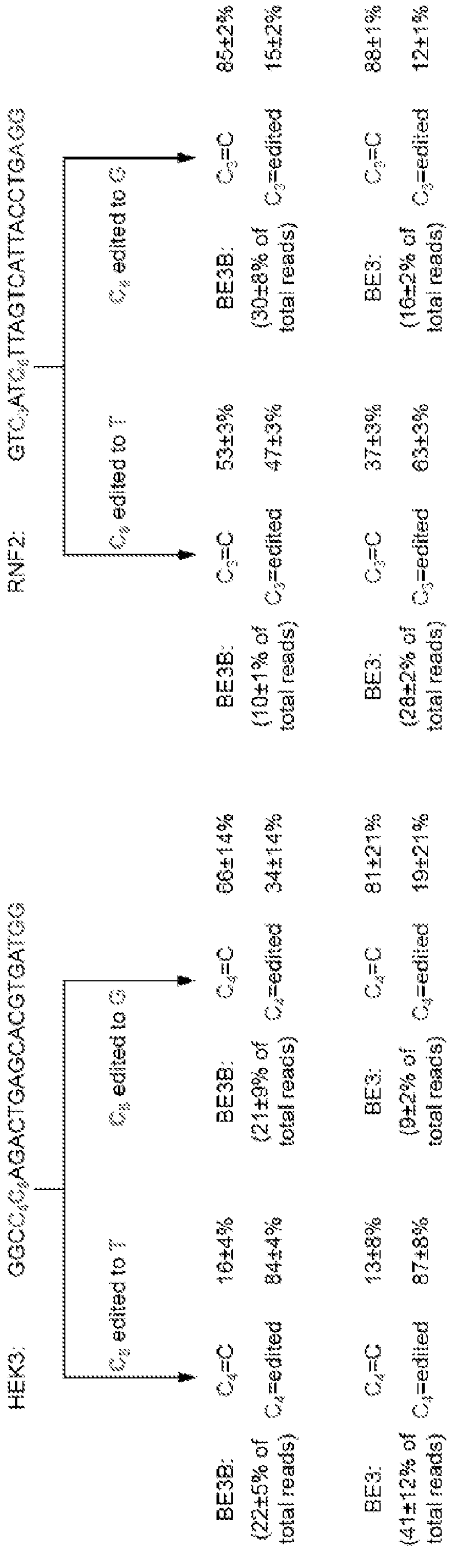

HEK3:
GGCC_4C_8AGACTGAGCACGTGATGG
C_8 edited to T
C_8 edited to G
BE3B: (22±5% of total reads)
C_4=C
16±4%
C_4=edited
84±4%
BE3B: (21±9% of total reads)
C_4=C
66±14%
C_4=edited
34±14%
BE3: (41±12% of total reads)
C_4=C
13±8%
C_4=edited
87±8%
BE3: (9±2% of total reads)
C_4=C
81±21%
C_4=edited
19±21%
RNF2:
GTC_3ATC_6TTAGTCATTACCTGAGG
C_6 edited to T
C_6 edited to G
BE3B: (10±1% of total reads)
C_3=C
53±3%
C_3=edited
47±3%
BE3B: (30±8% of total reads)
C_3=C
85±2%
C_3=edited
15±2%
BE3: (28±2% of total reads)
C_3=C
37±3%
C_3=edited
63±3%
BE3: (16±2% of total reads)
C_3=C
88±1%
C_3=edited
12±1%

FIGURE 143

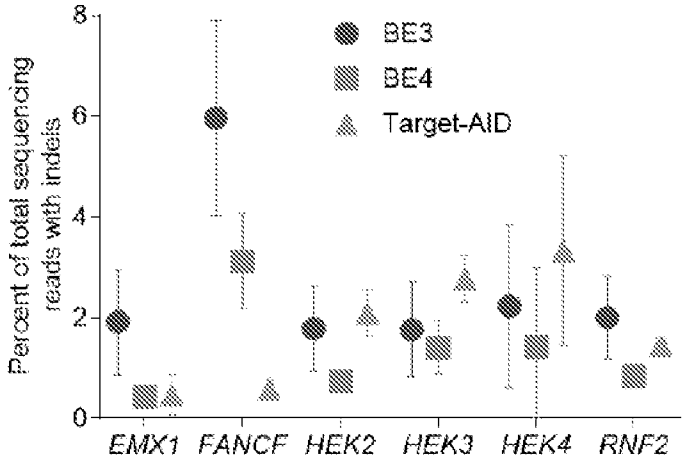


FIGURE 144A

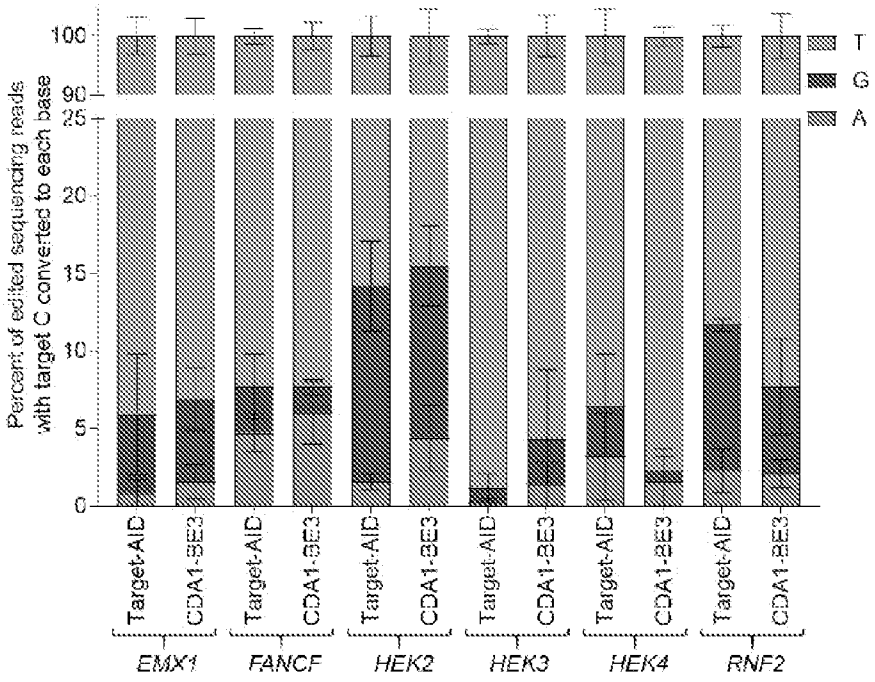


FIGURE 144B

EMX1: GAGTC$_5$CGAGCAGAAGAAGAAGGG

FANCF: GGAATCC$_7$CTTCTGCAGCACCTGG

HEK2: GAACAC$_5$AAAGCATAGACTGCGGG

HEK3: GGCCC$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTCATC$_5$TTAGTCATTACCTGAGG

FIGURE 144C

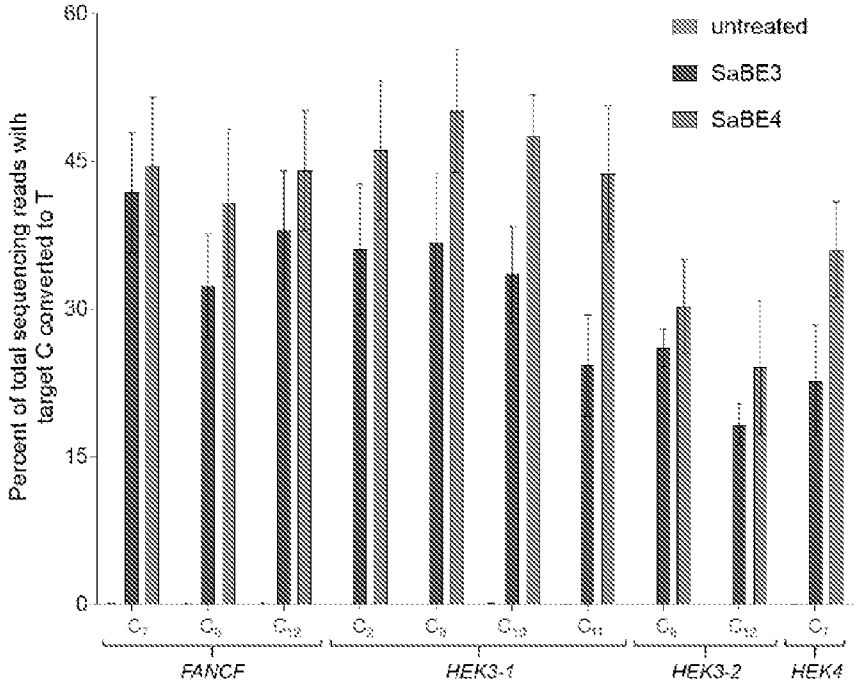


FIGURE 145A

FANCF: GATGTTC$_7$C$_9$AATC$_{12}$AGTACGCAGAGAGT

HEK3-1: TC$_2$TGCTTC$_5$TC$_{10}$C$_{11}$AGCCCTGGCCTGGGT

HEK3-2: ACGTGCTC$_5$AGTC$_{12}$TGGGCCCCAAGGAT

HEK4: GTGGCAC$_7$TGCGGCTGGAGGTGGGGGT

FIGURE 145B

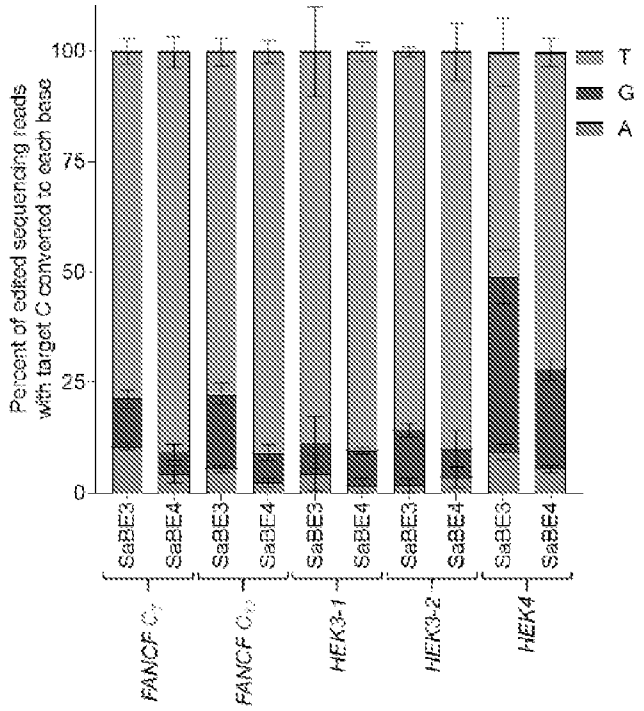
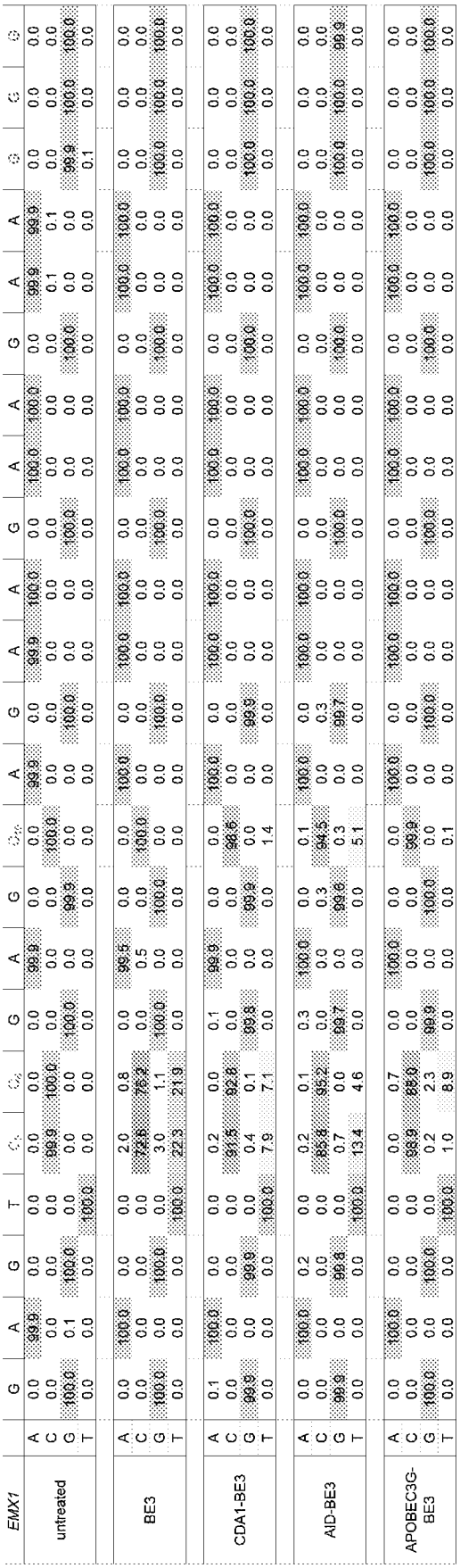


FIGURE 145C



EMX1		G	A	G	T			G	A	G		A	G	A	A	G	A	A	G	
untreated	A	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	100.0	100.0	0.0	99.9	99.9	0.0	0.0
	C	0.0	0.0	0.0	0.0	99.9	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0
	G	100.0	0.1	100.0	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	99.9	100.0
	T	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
BE3	A	0.0	100.0	0.0	0.0	2.0	0.8	0.0	99.5	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	73.6	76.2	0.0	0.5	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	100.0	0.0	3.0	1.1	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	100.0	22.3	21.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.1	100.0	0.0	0.0	0.2	0.0	0.1	99.9	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	91.5	92.8	0.0	0.0	0.0	96.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	0.0	99.9	0.0	0.4	0.1	99.8	0.0	99.9	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	100.0	7.9	7.1	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AID-BE3	A	0.0	100.0	0.2	0.0	0.2	0.1	0.3	100.0	0.0	0.1	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	85.8	95.2	0.0	0.0	0.3	94.5	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	0.0	99.8	0.0	0.7	0.0	99.7	0.0	99.6	0.3	0.0	99.7	0.0	0.0	100.0	0.0	0.0	100.0	99.9
	T	0.0	0.0	0.0	100.0	13.4	4.6	0.0	0.0	0.0	5.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
APOBEC3G-BE3	A	0.0	100.0	0.0	0.0	0.0	0.7	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	99.9	96.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	100.0	0.0	0.2	2.3	99.9	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	100.0	1.0	0.9	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

FIGURE 146

FANCF		G	G	A	A	T	C	C	T	T	C	T	G	C	A	G	A	C	C	T	G	G
untreated	A	0.1	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0
	G	99.9	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
BE3	A	0.0	0.1	100.0	100.0	0.0	1.0	1.4	1.4	0.0	0.0	0.2	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	76.7	75.0	74.4	0.0	0.0	64.8	0.0	0.0	99.9	0.0	0.0	100.0	99.9	0.0	0.0	0.0
	G	100.0	99.9	0.0	0.0	0.0	0.4	0.8	0.9	0.0	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0
	T	0.0	0.0	0.0	0.0	100.0	21.0	22.9	23.4	100.0	100.0	10.9	100.0	0.0	0.1	0.0	0.0	0.0	0.1	100.0	0.0	0.0
CDA1-BE3	A	0.1	0.1	100.0	100.0	0.0	0.3	0.6	0.3	0.0	0.0	0.0	0.0	0.1	0.0	100.0	0.1	0.0	100.0	0.0	0.0	0.1
	C	0.0	0.0	0.0	0.0	0.0	82.4	83.5	84.7	0.0	0.0	97.3	0.0	0.0	99.6	0.0	0.0	99.8	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.1	0.3	0.2	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	99.9	99.9
	T	0.0	0.0	0.0	0.0	100.0	17.2	15.6	14.7	99.9	100.0	2.6	100.0	0.0	0.3	0.0	0.0	0.2	0.0	0.0	0.0	0.0
AID-BE3	A	0.0	0.0	99.9	99.9	0.1	0.2	0.6	0.3	0.0	0.0	0.0	0.0	0.1	0.0	99.9	0.1	0.1	100.0	0.0	0.0	0.1
	C	0.0	0.0	0.0	0.0	0.0	73.8	80.3	86.6	0.0	0.0	96.7	0.0	0.0	98.8	0.0	0.0	97.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.8	0.5	0.4	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	99.9	99.9
	T	0.0	0.0	0.0	0.0	99.9	25.2	18.6	12.7	100.0	100.0	3.2	100.0	0.0	1.2	0.0	0.0	2.9	0.0	0.1	0.0	0.0
APOBEC3G-BE3	A	0.0	0.0	100.0	100.0	0.0	0.0	2.9	1.7	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	100.0	0.0	0.0	0.1
	C	0.0	0.0	0.0	0.0	0.0	96.1	89.3	86.3	0.0	0.0	99.8	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	G	100.0	99.9	0.0	0.0	0.0	0.3	1.6	10.8	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	99.9
	T	0.0	0.0	0.0	0.0	100.0	3.6	27.2	19.2	100.0	100.0	0.1	100.0	0.0	0.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0

FIGURE 147

HEK2		G	A	A	[illegible]	A	[illegible]	A	A	A	G	[illegible]	A	T	A	G	A	C	T	G	C	G	G
untreated	A	0.0	100.0	100.0	0.0	100.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	99.9	100.0	100.0
	T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0
BE3	A	0.1	100.0	100.0	1.3	99.9	0.7	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	80.5	0.0	46.5	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	100.0	0.0	0.0
	G	99.9	0.0	0.0	0.9	0.1	36.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	99.9	100.0
	T	0.0	0.0	0.0	17.2	0.0	14.8	0.0	0.0	0.0	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.2	100.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.0	99.9	99.9	9.6	99.9	1.8	100.0	100.0	100.0	0.0	0.2	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	70.3	0.0	73.6	0.0	0.0	0.0	0.0	90.7	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0
	G	100.0	0.0	0.1	0.3	0.1	3.5	0.0	0.0	0.0	100.0	0.3	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0
	T	0.0	0.0	0.0	19.8	0.0	21.2	0.0	0.0	0.0	0.0	8.8	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0
AID-BE3	A	0.0	99.9	99.9	6.9	99.9	1.2	100.0	100.0	100.0	0.0	0.2	100.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	24.5	0.0	22.4	0.0	0.0	0.0	0.0	69.8	0.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	99.8	0.0	0.0
	G	99.9	0.0	0.1	0.7	0.1	6.1	0.0	0.0	0.0	99.9	0.6	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.2	100.0	100.0
	T	0.1	0.1	0.0	27.0	0.0	40.3	0.0	0.0	0.0	0.1	15.5	0.0	100.0	0.0	0.0	0.2	99.9	0.0	0.0	0.2	0.0	0.0
APOBEC3G-BE3	A	0.0	100.0	100.0	0.1	100.0	0.1	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	97.8	0.0	80.3	0.0	0.0	0.0	0.0	98.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.1	0.0	5.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.1	100.0	100.0
	T	0.0	0.0	0.0	2.0	0.0	13.7	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0

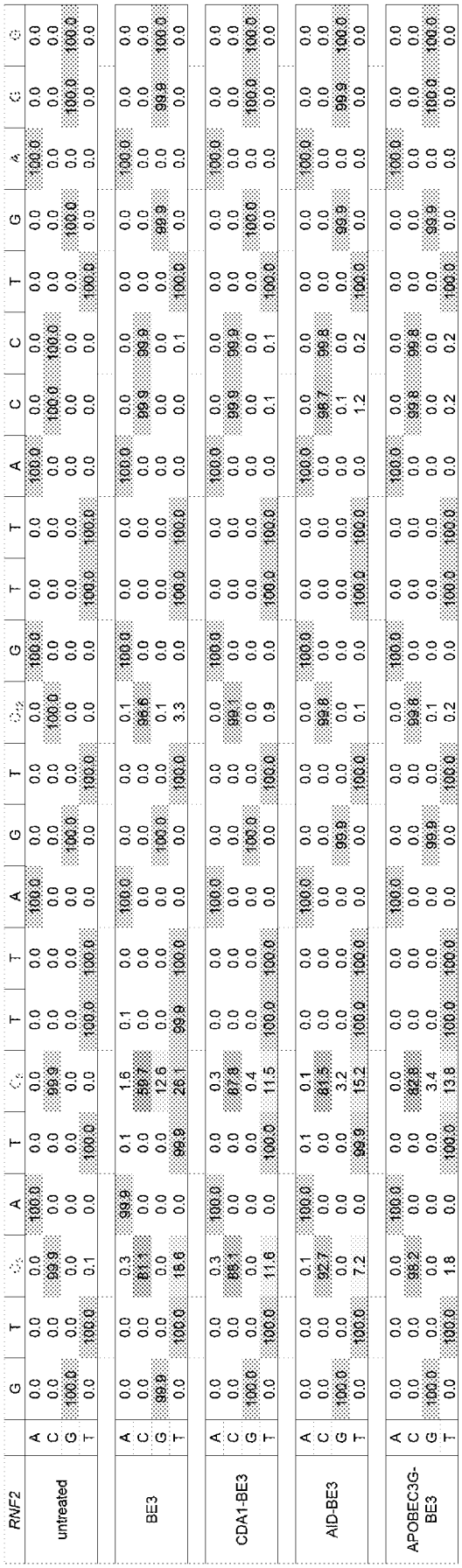
FIGURE 148

HEK3		G	G	C	C	A	G	A	C	T	G	A	G	C	A	C	T	G	A	T	C	C
untreated	A	0.0	0.1	0.0	0.0	99.9	0.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	99.9	0.0	0.0	0.1
	C	0.0	0.0	100.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
	G	100.0	99.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	99.9	0.0	99.9	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	99.9
	T	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0
BE3	A	0.0	0.0	0.0	0.8	1.1	100.0	0.0	100.0	0.4	0.0	0.0	100.0	0.2	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	99.6	67.4	98.1	0.0	0.0	0.0	94.3	0.0	0.0	0.0	100.0	0.0	99.7	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	100.0	0.0	0.4	8.9	0.0	100.0	0.0	0.6	0.0	100.0	99.8	0.0	0.0	0.0	100.0	0.0	100.0	0.0	99.9	100.0
	T	0.0	0.0	1.3	31.4	31.9	0.0	0.0	0.0	0.7	100.0	0.0	0.0	0.0	0.3	0.0	0.0	100.0	0.0	100.0	0.1	0.0
CDA1-BE3	A	0.1	0.0	0.8	0.6	0.0	100.0	0.0	99.9	0.4	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	99.9	0.0	0.1
	C	0.0	0.0	79.4	73.6	73.4	0.0	0.0	0.0	91.0	0.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.8	100.0	0.0	0.0	2.1	0.0	100.0	0.0	0.3	0.0	99.8	0.0	100.0	0.0	0.0	99.9	0.0	99.8	0.1	99.8	99.8
	T	0.1	0.0	19.8	25.9	24.4	0.0	0.0	0.0	8.3	100.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0
AID-BE3	A	0.1	0.3	1.0	0.7	0.3	99.9	0.0	100.0	1.6	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	83.9	75.4	80.2	0.0	0.0	0.0	86.6	0.0	0.0	0.0	98.9	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.8	99.7	0.1	0.4	1.0	0.0	100.0	0.0	1.2	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	99.9
	T	0.0	0.0	15.0	23.5	18.5	0.0	0.0	0.0	10.6	100.0	0.0	0.0	1.1	0.0	0.2	0.0	100.0	0.0	100.0	0.0	0.0
APOBEC3G-BE3	A	0.0	0.0	0.0	0.6	3.1	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	99.9	86.7	33.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	99.9	0.0	1.4	25.6	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	11.3	34.4	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0

FIGURE 149

HEK4		G	G	G	A	C	T	G	G	C	T	G	G	A	G	G	T	G	G	G	G
untreated	A	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	100.0	0.0	100.0	0.0	0.0	99.9	100.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	100.0	0.0	100.0	100.0	0.4	99.9	99.9	100.0	99.9
	T	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.5	0.1	0.0	0.0	0.0
BE3	A	0.1	0.0	0.1	99.9	8.4	0.1	0.0	0.1	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	99.0	0.0	41.9	0.0	0.0	99.3	99.8	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	24.5	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.1	100.0	100.0	0.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	0.9	0.0	25.2	99.9	0.0	0.5	0.2	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.2	0.1	0.5	100.0	0.5	0.0	0.0	0.2	0.1	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	91.3	0.0	97.5	0.0	0.0	93.4	97.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.8	99.9	0.3	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	7.8	0.0	12.0	100.0	0.0	6.3	2.9	100.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0
AID-BE3	A	0.3	0.4	0.3	100.0	0.9	0.0	0.0	0.9	0.1	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	81.5	0.0	71.3	0.0	0.0	80.7	97.9	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0
	G	99.7	99.6	0.1	0.0	0.7	0.0	100.0	0.1	0.0	0.0	99.9	100.0	0.0	100.0	100.0	0.2	99.9	100.0	100.0	99.9
	T	0.0	0.0	18.0	0.0	27.1	100.0	0.0	18.3	2.0	99.9	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.1
APOBEC3G-BE3	A	0.1	0.0	0.0	99.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	99.9	0.0	99.4	0.0	0.0	99.9	100.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.4	0.0	99.9	0.0	0.0	0.0	99.9	100.0	0.1	100.0	99.9	0.0	99.9	100.0	100.0	99.9
	T	0.0	0.0	0.0	0.0	0.2	99.9	0.0	0.1	0.0	99.9	0.1	0.0	0.0	0.0	0.1	99.9	0.0	0.0	0.0	0.1

FIGURE 150



RNF2		G	T		A	T		T	A	G	T		T	A	C	C	T	G		
untreated	A	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0
	C	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0
	T	0.0	100.0	0.1	0.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
BE3	A	0.0	0.0	0.3	99.9	0.1	1.6	0.1	100.0	0.0	0.0	0.1	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	C	0.0	0.0	81.4	0.0	0.0	56.7	0.0	0.0	0.0	0.0	96.6	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0
	G	99.9	0.0	0.0	0.0	0.0	12.6	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	99.9	0.0	99.9
	T	0.0	100.0	18.6	0.0	99.9	26.1	99.9	100.0	0.0	100.0	3.3	0.0	100.0	100.0	0.1	100.0	0.0	0.0	0.0
CDA1-BE3	A	0.0	0.0	0.3	100.0	0.0	0.3	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	C	0.0	0.0	88.1	0.0	0.0	87.8	0.0	0.0	0.0	0.0	99.1	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0
	T	0.0	100.0	11.8	0.0	100.0	11.5	100.0	100.0	0.0	100.0	0.9	0.0	100.0	100.0	0.1	0.1	0.0	0.0	0.0
AID-BE3	A	0.0	0.0	0.1	100.0	0.1	0.1	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	C	0.0	0.0	92.7	0.0	0.0	81.5	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	98.7	99.8	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	3.2	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.1	0.0	99.9	0.0	99.9
	T	0.0	100.0	7.2	0.0	99.9	15.2	100.0	100.0	0.0	100.0	0.1	0.0	100.0	100.0	0.0	0.2	100.0	0.0	0.0
APOBEC3G-BE3	A	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	C	0.0	0.0	96.2	0.0	0.0	92.8	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	99.8	99.8	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	3.4	0.0	0.0	99.9	0.0	0.1	0.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0
	T	0.0	100.0	1.8	0.0	100.0	13.8	100.0	0.0	0.0	100.0	0.2	0.0	100.0	100.0	0.2	0.2	100.0	0.0	0.0

FIGURE 151

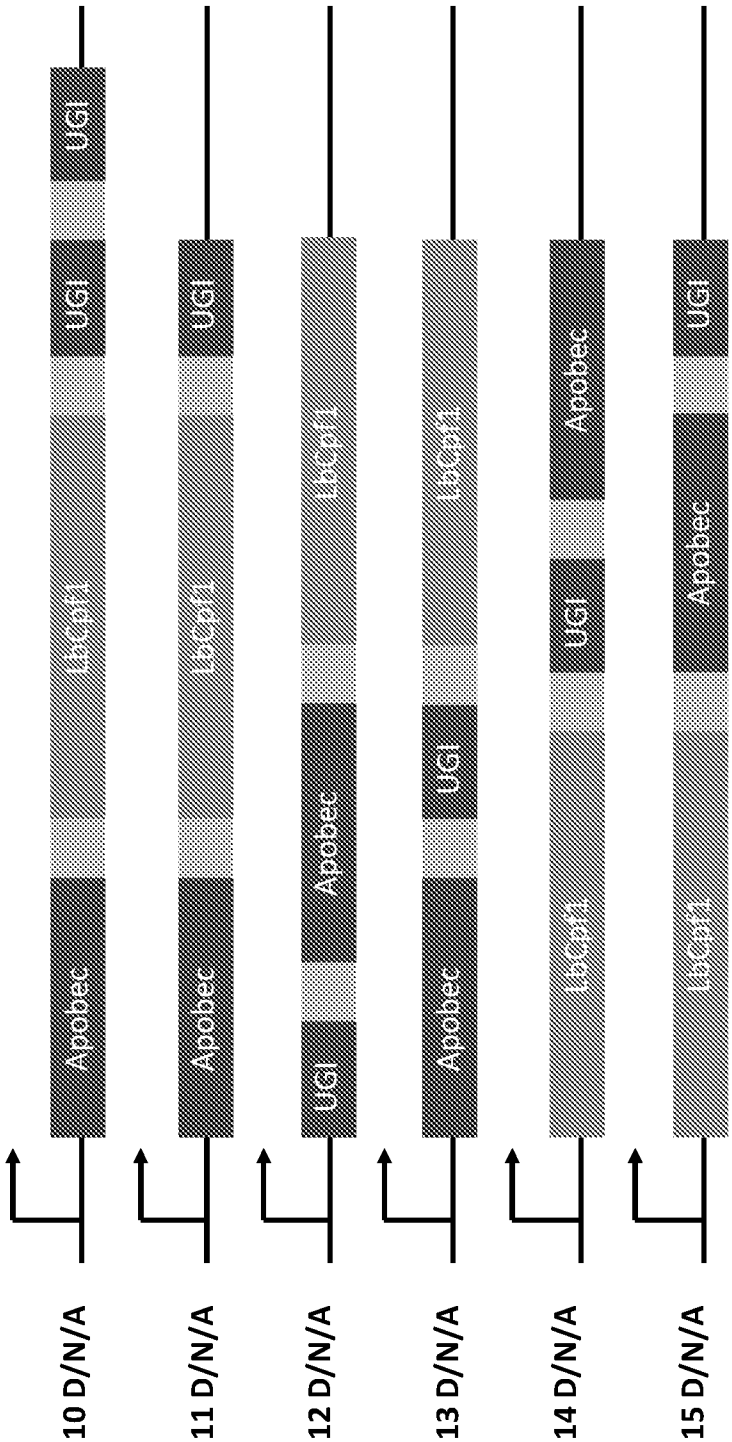


FIGURE 152

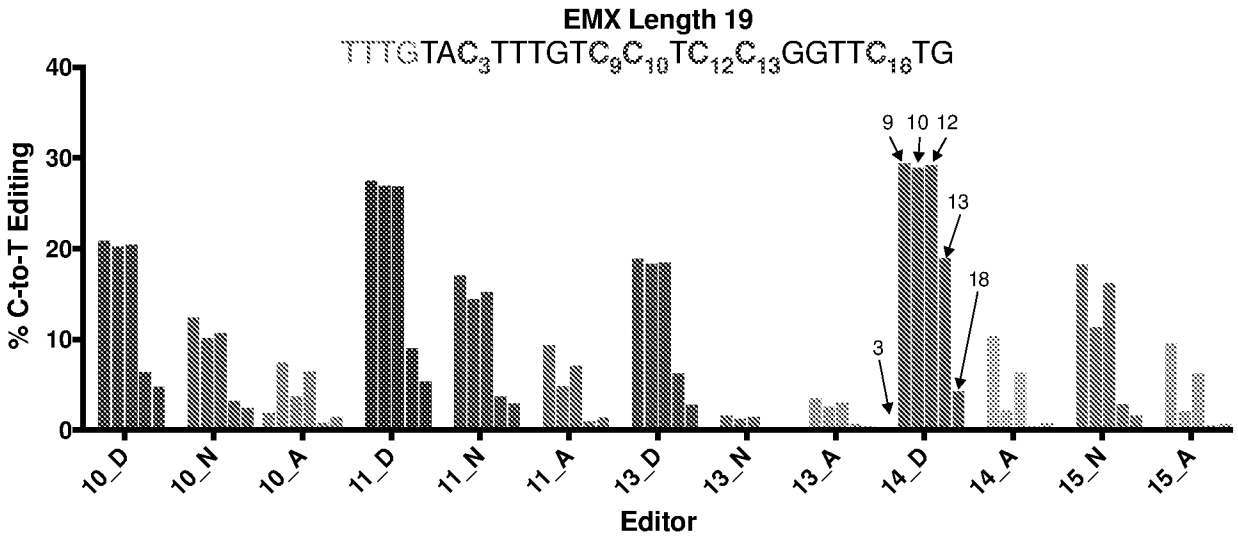


FIGURE 153

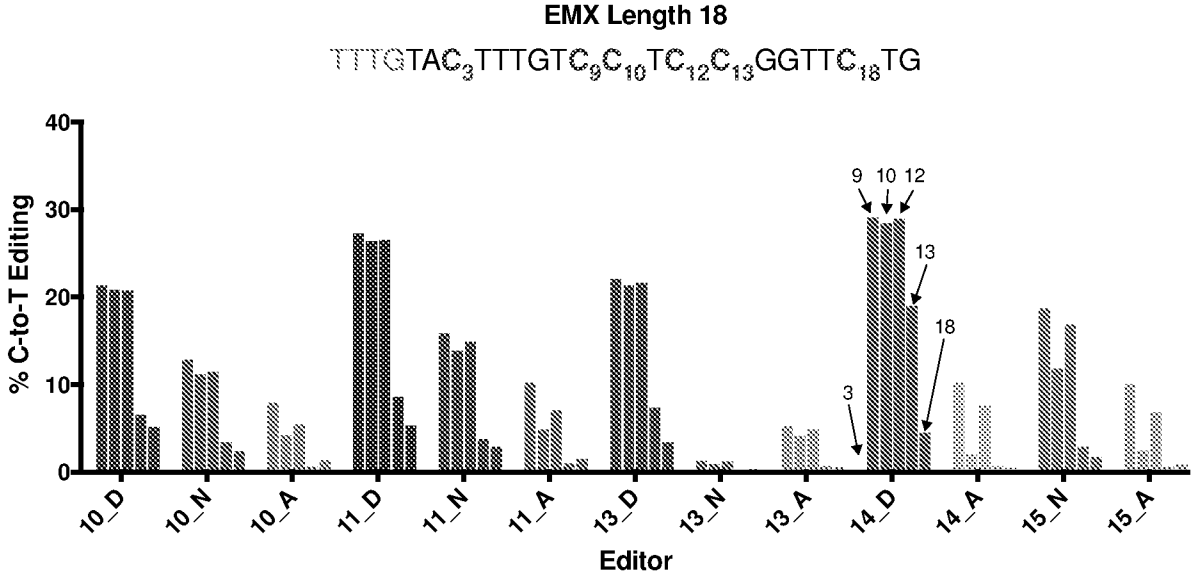


FIGURE 154

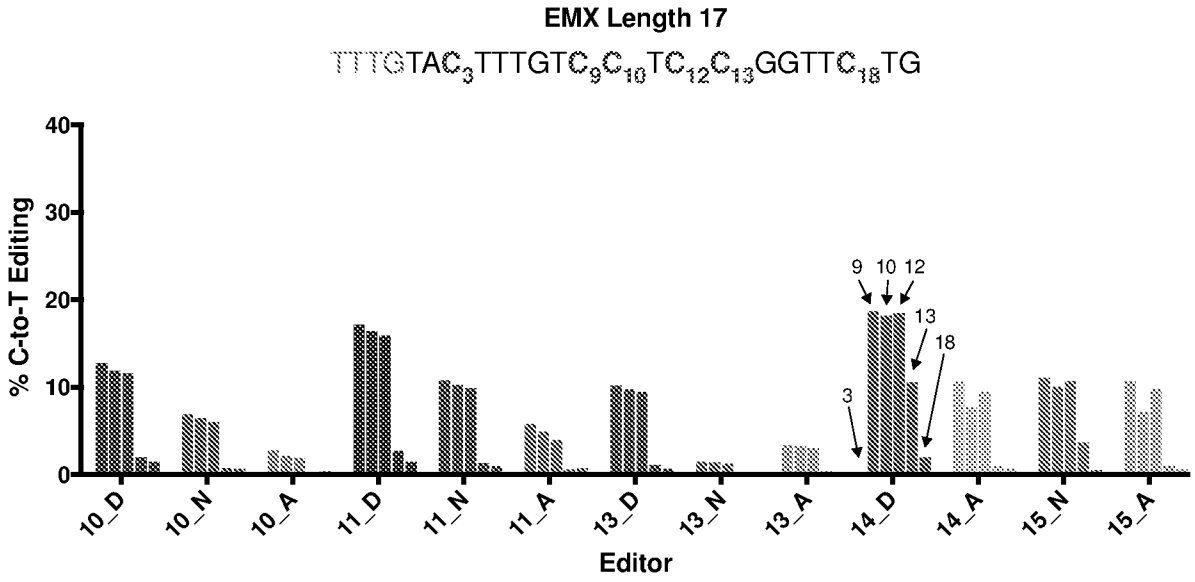


FIGURE 155

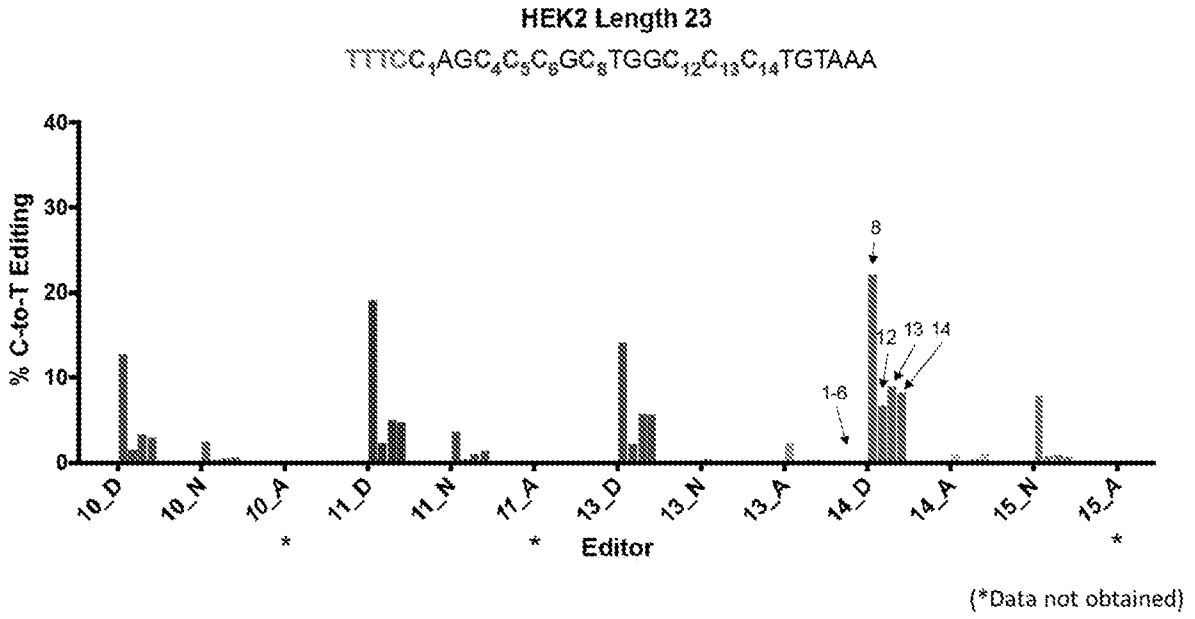


FIGURE 156

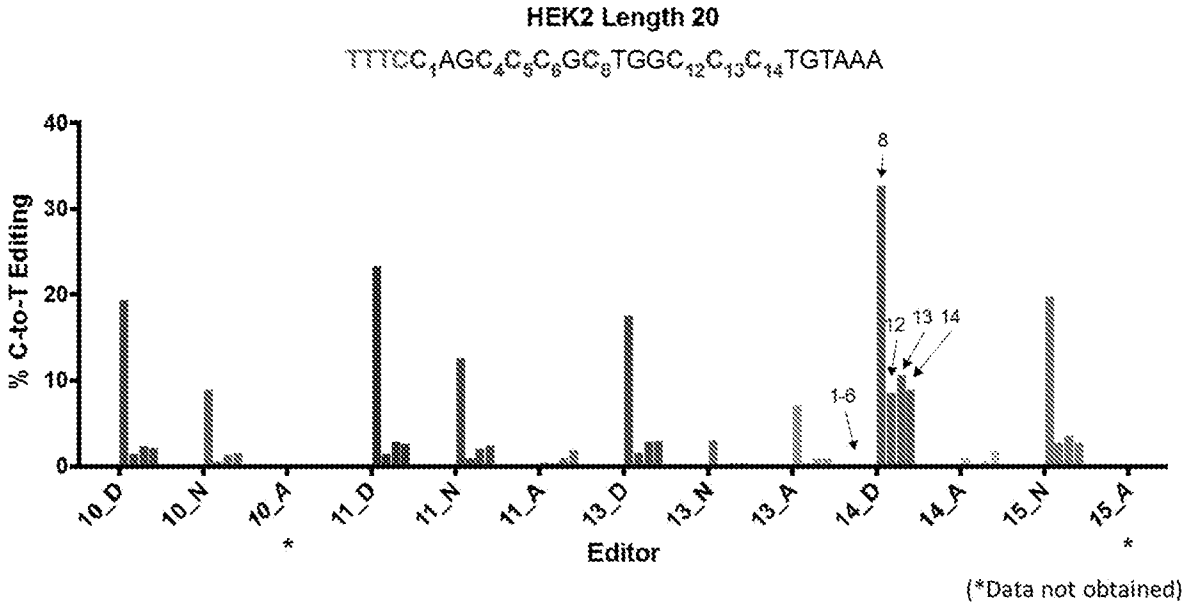

HEK2 Length 20
TTTCC1AGC4C5C6GC8TGGC12C13C14TGTAAA
% C-to-T Editing
40
30
20
10
0
8
12
13
14
1-6
10_D
10_N
10_A
11_D
11_N
11_A
13_D
13_N
13_A
14_D
14_A
15_N
15_A
*
*
Editor
(*Data not obtained)

FIGURE 157

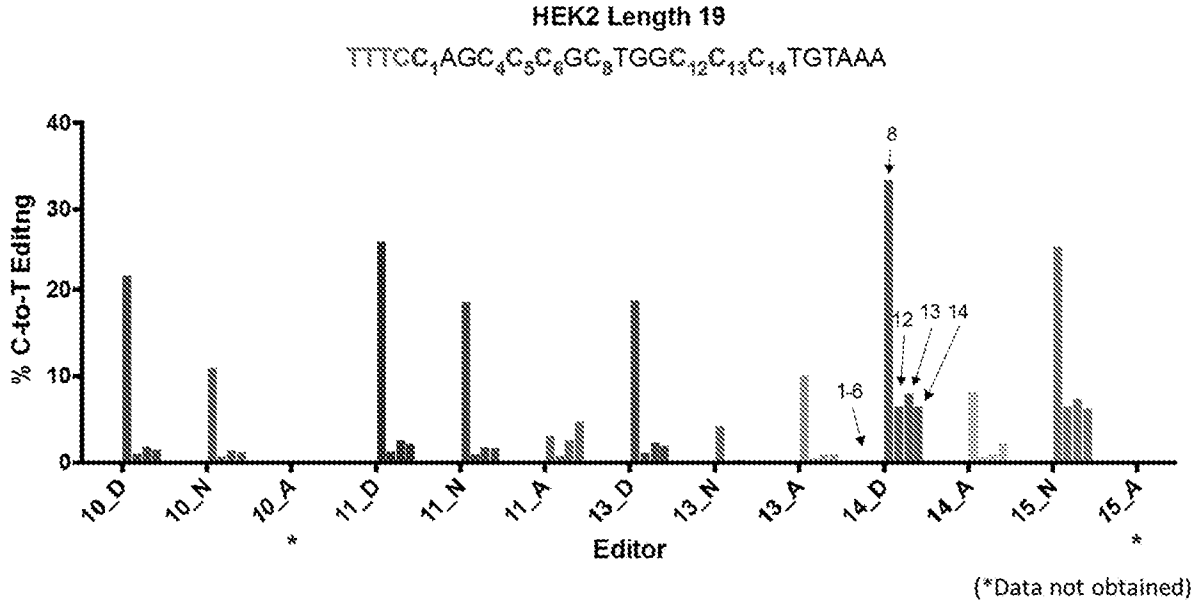


FIGURE 158

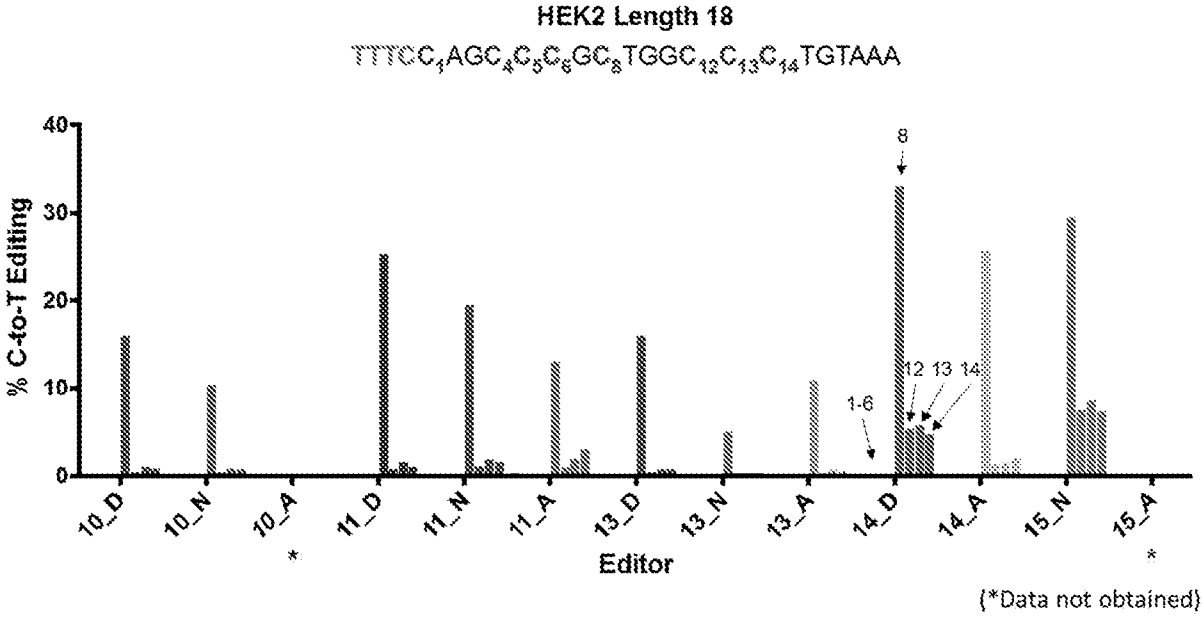


FIGURE 159

EMX 19

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.025	0.039	2.662	0.012	0.013	0.041	0.02	0.013	0.012	0.033	0.012	0.02	0.021
C9	20.899	13.15	10.456	27.535	18.921	13.072	19.043	1.591	3.638	29.39	14.271	20.95	13.924
C10	20.226	10.784	5.206	26.909	16.026	6.748	18.445	1.266	2.658	28.933	2.934	13.039	3.051
C12	20.453	11.385	9.095	26.849	16.893	9.983	18.589	1.437	3.127	29.165	8.695	18.597	9.07
C13	6.39	3.396	1.085	8.946	4.126	1.3	6.317	0.257	0.659	18.973	0.547	3.247	0.8
C18	4.757	2.603	2.013	5.304	3.233	1.914	2.805	0.235	0.367	4.302	0.992	1.848	0.994
indels (%)	0.028979	6.0862	28.9183	0.038623	9.8557	28.6068	0.94296	1.4573	3.78	0.024623	27.6872	13.0741	31.3295
C3	0.02499276	0.03662638	1.89219485	0.01199537	0.01171876	0.02927121	0.01981141	0.01281055	0.0115464	0.03299187	0.00867754	0.01738518	0.01442081
C9	20.8929437	12.3496647	7.43230255	27.5243652	17.056203	9.3325191	18.8634321	1.56781436	3.5004836	29.3827633	10.3197597	18.2109761	9.56168042
C10	20.2201387	10.1276642	3.7005133	26.8986069	14.4465255	4.81761314	18.271071	1.24755058	2.5575276	28.9258758	2.12165755	11.3342681	2.09513696
C12	20.4470729	10.6920861	6.46488062	26.8386301	15.2280766	7.12718316	18.4137132	1.4160586	3.0087994	29.1578187	6.28759796	16.1656096	6.22841435
C13	6.38814824	3.18931265	0.77123645	8.94254479	3.71935382	0.9281116	6.25743322	0.25325474	0.6340898	18.9683283	0.39555102	2.82248397	0.549364
C18	4.75562147	2.44457621	1.43087462	5.30195144	2.91436522	1.36646585	2.77854997	0.23157535	0.3531274	4.30094072	0.71734298	1.60639063	0.68258477

FIGURE 160

EMX 18

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.023	0.008	0.032	0.044	0.035	0.016	0.03	0.016	0.016	0.018	0.02	0.017	0.01
C9	21.333	13.606	10.333	27.294	17.216	14.646	22.315	1.297	5.485	29.094	14.566	21.847	13.66
C10	20.865	11.831	5.528	26.409	15.041	6.931	21.548	0.932	4.309	28.456	2.996	13.762	3.337
C12	20.773	12.165	7.147	26.572	16.178	10.157	21.839	1.202	5.059	28.937	10.88	19.625	9.359
C13	6.538	3.591	0.848	8.562	4.096	1.387	7.426	0.129	0.729	18.993	0.945	3.353	0.809
C18	5.174	2.54	1.788	5.313	3.146	2.196	3.443	0.243	0.565	4.517	0.647	2.03	1.154
indels (%)	0.039639	5.7886	23.1911	0.035862	7.9573	30.2671	1.1014	1.1121	4.6798	0.040521	29.8821	14.2217	26.5552
C3	0.02299088	0.00753691	0.02457885	0.04398422	0.03221495	0.01115726	0.02966958	0.01582206	0.01525123	0.01799271	0.01402358	0.01458231	0.00734448
C9	21.3245438	12.8184031	7.93666364	27.2842118	15.8460712	10.2130805	22.0692226	1.28257606	5.22831297	29.0822108	10.2133733	18.7399852	10.0325597
C10	20.8567293	11.1461507	4.24599599	26.3995292	13.8441425	4.8331873	21.3106703	0.92163523	4.10734742	28.4444693	2.10073228	11.8048096	2.45085298
C12	20.7647658	11.4608168	5.48953208	26.5624707	14.890668	7.08277065	21.5984653	1.18863256	4.82224892	28.9252744	7.62882752	16.8339914	6.87369883
C13	6.5354084	3.38313137	0.65133947	8.5589295	3.77006899	0.96719532	7.34421004	0.12756539	0.69488426	18.9853038	0.66261416	2.8761464	0.59416843
C18	5.17194908	2.39296956	1.37334313	5.31109465	2.89566334	1.53133448	3.4050788	0.2402976	0.53855913	4.51516967	0.45366281	1.74129949	0.84755299

FIGURE 161

EMX 17

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.016	0.014	0.044	0.012	0.018	0.012	0.013	0.016	0.03	0.014	0.011	0.016	0.005
C9	12.742	6.895	2.862	17.142	10.77	5.941	10.177	1.471	3.395	18.719	11.13	11.103	11.493
C10	11.901	6.427	2.199	16.385	10.277	5.064	9.741	1.393	3.216	18.206	8.1	10.086	7.747
C12	11.548	6.014	1.916	15.933	9.872	4.041	9.468	1.287	2.986	18.492	9.847	10.751	10.523
C13	1.993	0.768	0.252	2.696	1.357	0.599	1.155	0.083	0.285	10.532	0.971	3.674	1.082
C18	1.457	0.613	0.361	1.471	0.987	0.773	0.667	0.043	0.245	1.978	0.729	0.545	0.621
indels (%)	0.017245	0.052423	1.7787	0.025203	0.11499	2.7876	0.14938	0.019738	0.041079	0.04713	4.2923	0.37931	6.7995
C3	0.01599724	0.01399266	0.04321737	0.01199698	0.0179793	0.01166549	0.01298058	0.01599684	0.02998768	0.0139934	0.01052785	0.01593931	0.00466003
C9	12.7398026	6.89138543	2.81109361	17.1376797	10.7576156	5.77538868	10.1617976	1.47070965	3.39360537	18.7101777	10.652267	11.0608852	10.7115335
C10	11.8989477	6.42363077	2.15988639	16.3808705	10.2651825	4.92283594	9.72644889	1.39272505	3.2146789	18.1974195	7.7523237	10.0477428	7.22024274
C12	11.5460085	6.01084728	1.88192011	15.9289844	9.86064819	3.92835308	9.4538567	1.28674597	2.98477338	18.4832847	9.42433722	10.7102204	9.80748862
C13	1.99265631	0.76759739	0.24751768	2.69532053	1.35543959	0.58230228	1.15327466	0.08298362	0.28488292	10.5270363	0.92932177	3.66006415	1.00842941
C18	1.45674874	0.61267865	0.35457889	1.47062926	0.98586505	0.75145185	0.66600364	0.04299151	0.24489936	1.97706777	0.69770913	0.54293276	0.57877511

FIGURE 162

Hek2 23

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.042	0.021	0	0.039	0.084	0	0.045	0.07	0.04	0.045	0.056	0.057	0
C4	0.016	0.018	0	0.023	0.013	0	0.027	0.022	0.016	0.035	0.008	0.016	0
C5	0.031	0.018	0	0.037	0.052	0	0.048	0.021	0.044	0.047	0.157	0.017	0
C6	0.064	0.045	0	0.064	0.02	0	0.077	0.068	0.051	0.078	0.032	0.08	0
C8	12.758	2.444	0	19.198	3.66	0	14.134	0.423	2.22	22.158	0.977	7.763	0
C12	1.418	0.291	0	2.281	0.446	0	2.163	0.101	0.21	6.745	0.235	0.809	0
C13	3.282	0.603	0	5.021	1.121	0	5.713	0.042	0.308	8.933	0.599	0.921	0
C14	2.898	0.641	0	4.719	1.38	0	5.614	0.096	0.38	8.268	1.066	0.746	0
indels (%)	0.052343	11.0664		0.041898			2.2549	6.3138	13.9681	0.085777	59.2508	11.3052	
C1	0.04197802	0.01867606	0	0.03898366	0.084	0	0.0439853	0.06558034	0.03441276	0.0449614	0.02281955	0.05055604	0
C4	0.01599163	0.01600805	0	0.02299036	0.013	0	0.02639118	0.02061096	0.0137651	0.03496998	0.00325994	0.01419117	0
C5	0.03098377	0.01600805	0	0.0369845	0.052	0	0.04691765	0.0196741	0.03785404	0.04695968	0.06397624	0.01507812	0
C6	0.0639665	0.04002012	0	0.06397319	0.02	0	0.07526373	0.06370662	0.04387627	0.07793309	0.01303974	0.07095584	0
C8	12.7513221	2.17353718	0	19.1899564	3.66	0	13.8152924	0.39629263	1.90990818	22.1389935	0.39811968	6.88537732	0
C12	1.41725778	0.25879678	0	2.28004431	0.446	0	2.11422651	0.09462306	0.18066699	6.73921434	0.09576062	0.71754093	0
C13	3.2802821	0.53626961	0	5.0188963	1.121	0	5.58417756	0.0393482	0.26497825	8.92533754	0.24408771	0.81687911	0
C14	2.8964831	0.57006438	0	4.71702283	1.38	0	5.48740991	0.08993875	0.32692122	8.26090796	0.43438647	0.66166321	0

FIGURE 163

Hek2 20

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.024	0.051	0	0.036	0.116	0.126	0.035	0.022	0.042	0.032	0.02	0.133	0
C4	0.054	0.023	0	0.026	0.011	0.01	0.019	0.019	0.011	0.025	0.016	0.016	0
C5	0.033	0.017	0	0.051	0.054	0.048	0.046	0.02	0.029	0.035	0.075	0.043	0
C6	0.096	0.083	0	0.058	0.032	0.022	0.055	0.016	0.047	0.139	0.166	0.083	0
C8	19.339	9.265	0	23.342	13.199	1.173	17.974	3.225	7.353	32.768	2.835	20.891	0
C12	1.402	0.551	0	1.491	1.036	0.713	1.68	0.188	0.471	8.549	0.3	2.967	0
C13	2.377	1.409	0	2.974	2.174	2.814	3.043	0.356	0.895	10.654	1.81	3.81	0
C14	2.129	1.545	0	2.604	2.523	5.582	3.076	0.316	0.955	8.964	4.97	3.086	0
indels (%)	0.10237	3.2815	24.3902	0.10375	4.3674	66.4779	1.8097	1.6132	3.0719	0.093678	63.4082	5.4543	
C1	0.02397543	0.04932644	0	0.03596265	0.11093382	0.04223785	0.03436661	0.0216451	0.0407098	0.03197002	0.00731836	0.12574578	0
C4	0.05394472	0.02224526	0	0.02597303	0.01051959	0.00335221	0.01865616	0.01869349	0.01066209	0.02497658	0.00585469	0.01512731	0
C5	0.03296622	0.01644215	0	0.05094709	0.0516416	0.01609061	0.04516754	0.01967736	0.02810915	0.03496721	0.02744385	0.04065465	0
C6	0.09590172	0.08027636	0	0.05793983	0.03060243	0.00737486	0.05400467	0.01574189	0.04555621	0.13886979	0.06074239	0.07847293	0
C8	19.3192027	8.96096903	0	23.3177827	12.6225469	0.39321423	17.6487245	3.1729743	7.12712319	32.7373036	1.03737753	19.7515422	0
C12	1.40056477	0.53291894	0	1.48945309	0.99075374	0.23901257	1.64959704	0.18496718	0.45653135	8.54099147	0.1097754	2.80517092	0
C13	2.37456667	1.36276367	0	2.97091448	2.07905272	0.94331189	2.98793083	0.35025701	0.8675065	10.6440195	0.66231158	3.60219117	0
C14	2.12682054	1.49430083	0	2.60129835	2.4128105	1.87120362	3.02033363	0.31090229	0.92566336	8.9556027	1.81861246	2.9176803	0

FIGURE 164

Hek2 19

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.072	0.031	0	0.039	0.062	0.406	0.023	0.019	0.025	0.06	0.224	0.04	0
C4	0.034	0.019	0	0.021	0.06	0.018	0.039	0.033	0.025	0.012	0.009	0.008	0
C5	0.03	0.037	0	0.044	0.045	0.027	0.018	0.031	0.033	0.024	0.021	0.012	0
C6	0.053	0.066	0	0.042	0.071	0.164	0.033	0.031	0.046	0.098	0.117	0.034	0
C8	21.611	11.001	0	25.556	18.67	6.38	18.944	4.272	10.143	33.441	15.227	25.017	0
C12	1.049	0.704	0	1.345	1.022	1.548	1.154	0.195	0.489	6.54	1.177	6.539	0
C13	1.914	1.39	0	2.629	1.819	5.23	2.301	0.253	0.881	8.065	1.823	7.452	0
C14	1.545	1.212	0	2.238	1.761	9.623	1.98	0.151	0.975	6.517	4.174	6.365	0
indels (%)	0.056704	0.17936		0.04541	0.43984	50.2032	0.61981	0.18691	0.22381	0.14694	45.816	0.52493	
C1	0.07195917	0.0309444	0	0.03898229	0.0617273	0.20217501	0.02285744	0.01896449	0.02494405	0.05991184	0.12137216	0.03979003	0
C4	0.03398072	0.01896592	0	0.02099046	0.0597361	0.00896342	0.03875827	0.03293832	0.02494405	0.01198237	0.00487656	0.00795801	0
C5	0.02998299	0.03693364	0	0.04398002	0.04480207	0.01344514	0.01788843	0.03094206	0.03292614	0.02396473	0.01137864	0.01193701	0
C6	0.05296995	0.06588162	0	0.04198093	0.07068771	0.08166675	0.03279546	0.03094206	0.04589705	0.097856	0.06339528	0.03382152	0
C8	21.5987457	10.9812686	0	25.544395	18.5878819	3.17703584	18.8265832	4.2640152	10.120299	33.3918618	8.25059768	24.8856783	0
C12	1.04840518	0.70273731	0	1.34438924	1.01750484	0.77085446	1.14684739	0.19463553	0.48790557	6.53039012	0.63774568	6.50467483	0
C13	1.91291469	1.3875069	0	2.62780617	1.81099931	2.60437264	2.28673817	0.25252712	0.87902823	8.05314929	0.98777432	7.41288222	0
C14	1.54412392	1.20982616	0	2.23698372	1.75325442	4.79194606	1.96772776	0.15071777	0.97281785	6.50742392	2.26164016	6.33158821	0

FIGURE 165

Hek2 18

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.037	0.027	0	0.035	0.041	0.315	0.039	0.034	0.034	0.028	0.125	0.074	0
C4	0.014	0.016	0	0.007	0.018	0.017	0.018	0.038	0.019	0.014	0.021	0.035	0
C5	0.052	0.018	0	0.051	0.034	0.045	0.033	0.039	0.022	0.014	0.088	0.035	0
C6	0.056	0.035	0	0.032	0.08	0.047	0.082	0.051	0.07	0.081	0.208	0.107	0
C8	16.035	10.427	0	25.298	19.509	14.808	15.991	5.043	10.926	33.064	28.458	29.702	0
C12	0.571	0.503	0	0.802	1.161	1.138	0.513	0.228	0.317	5.485	1.471	7.611	0
C13	1.12	0.903	0	1.607	1.869	2.299	0.803	0.315	0.709	5.902	1.741	8.752	0
C14	0.877	0.819	0	1.132	1.609	3.468	0.751	0.375	0.636	4.806	2.316	7.414	0
indels (%)	0.076858	0.17892		0.048785	0.13947	11.5608	0.14868	0.1177	0.076621	0.23744	9.8678	0.50485	
C1	0.03697156	0.02695169	0	0.03498293	0.04094282	0.27858348	0.03894201	0.03395998	0.03397395	0.02793352	0.11266525	0.07362641	0
C4	0.01398924	0.01597137	0	0.00699659	0.0179749	0.01503466	0.01797324	0.03795527	0.01898544	0.01396676	0.01892776	0.0348233	0
C5	0.05196003	0.01796779	0	0.05097512	0.03395258	0.03979764	0.03295094	0.0389541	0.02198314	0.01396676	0.07931634	0.0348233	0
C6	0.05595696	0.03493738	0	0.03198439	0.07988842	0.04156642	0.08187808	0.05093997	0.06994637	0.08080767	0.18747498	0.10645981	0
C8	16.0226758	10.408344	0	25.2856584	19.4817908	13.0960767	15.9672246	5.03706439	10.9176284	32.9854928	25.6498215	29.5520495	0
C12	0.57056114	0.50210003	0	0.80160874	1.15938075	1.0064381	0.51223727	0.22773164	0.31675711	5.47197642	1.32584466	7.57257587	0
C13	1.11913919	0.90138435	0	1.60621603	1.86639331	2.03321721	0.8018061	0.31462925	0.70845676	5.88798629	1.5692016	8.70781553	0
C14	0.87632596	0.81753465	0	1.13144775	1.60675593	3.06707146	0.74988341	0.37455863	0.63551269	4.79458863	2.08746175	7.37657042	0

FIGURE 166

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/024208

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/11 C12N15/10
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal , BIOSIS, WPI Data, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ALEXIS C. KOMOR ET AL: "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage", NATURE, vol. 533, no. 7603, 20 April 2016 (2016-04-20), pages 420-424, XP055343871, GB ISSN: 0028-0836, DOI: 10.1038/nature17946 the whole document -/--	1,3-26, 31-90, 114-117, 127-136, 176-190, 192-207

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) orwhich is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 June 2018	23/08/2018

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Macchia, Giovanni

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/024208

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	-& ALEXIS C. KOMOR ET AL: "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage", NATURE, vol. 533, no. 7603, 20 April 2016 (2016-04-20), pages 420-424, XP055481330, GB ISSN: 0Q28-Q836, D0I: 10.1038/naturel7946 the whole document	1,3-26, 31-90, 114-117, 127-136, 176-190, 192-207
A	PL0SKY BRIAN S.: "CRISPR-mediated base editing without DNA double-strand breaks", MOLECULAR CELL, vol. 62, no. 4, 19 May 2016 (2016-05-19), pages 477-478, XP029552452, ISSN: 1097-2765, D0I: 10.1016/J.M0LCEL.2016.05.006 the whole document	176
Y	SERGEY SHMAK0V ET AL.: "Discovery and functional characterization of diverse class 2 CRISPR-Cas systems", MOLECULAR CELL, vol. 60, no. 3, 1 November 2015 (2015-11-01), pages 385-397, XP055482679, AMSTERDAM, NL ISSN: 1097-2765, D0I: 10.1016/j.molcel.2015.10.008 the whole document	1,8-19, 23-26, 31-90, 114-117, 127-136, 176-190, 192-207
A	LEWIS KEVIN M. AND AI LONG KE: "Building the class 2 CRISPR-Cas arsenal", MOLECULAR CELL, ELSEVIER, AMSTERDAM, NL, vol. 65, no. 3, 2 February 2017 (2017-02-02), pages 377-379, XP029906286, ISSN: 1097-2765, D0I: 10.1016/J.MOLCEL.2017.01.024 the whole document	8-19
Y	DAVID BURSTEIN ET AL.: "New CRISPR-Cas systems from uncultivated microbes", NATURE, vol. 542, no. 7640, 22 December 2016 (2016-12-22), pages 237-241, XP055480893, GB ISSN: 0028-0836, D0I: 10.1038/nature21059 the whole document	1,3-7, 23-26, 31-90, 114-117, 127-136, 176-190, 192-207

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/024208

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HUI YANG AND DINSHAW J. PATEL: "New CRISPR-Cas systems discovered" , CELL RESEARCH - XIBAO YANJIU , vol . 27, no. 3 , 21 February 2017 (2017-02-21) , pages 313-314, XP055481126, GB, CN ISSN : 1001-0602, DOI : 10. 1038/cr.2017 .21 the whole document	3-7

Y	EMINE KAYA ET AL. : "A bacterial Argonaute with noncanonical guide RNA specificity" , PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol . 113 , no. 15 , 30 March 2016 (2016-03-30) , pages 4057-4062 , XP055387813, US ISSN : 0027-8424, DOI : 10. 1073/pnas . 1524385113 the whole document	1 ,20-26, 31-90, 114-117, 127-136, 176-190, 192-207
Y	-& EMINE KAYA ET AL: "A bacterial Argonaute with noncanonical guide RNA specificity" , PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol . 113 , no. 15 , 30 March 2016 (2016-03-30) , pages 4057-4062 , XP055482683, US ISSN : 0027-8424, DOI : 10. 1073/pnas . 1524385113 the whole document	1 ,20-26, 31-90, 114-117, 127-136, 176-190, 192-207

A	K0M0R ALEXIS C. ET AL. : "CRISPR-based technologies for the manipulation of eukaryotic genomes" , CELL, CELL PRESS, AMSTERDAM, NL, vol . 168, no. 1 , 17 November 2016 (2016-11-17) , pages 20-36, XP029882172, ISSN : 0092-8674, DOI : 10. 1016/J . CELL.2O16. 10.O44 the whole document	1

A	Y BILL KIM ET AL. : "Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions" , NATURE BIOTECHNOLOGY, vol . 35, no. 4 , 13 February 2017 (2017-02-13) , pages 371-376, XP055481436, ISSN : 1087-0156, DOI : 10. 1038/nbt.3803 the whole document	1

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/024208

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	-& Y BILL KIM ET AL: "Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions", NATURE BIOTECHNOLOGY, vol. 35, no. 4, 13 February 2017 (2017-02-13), pages 371-376, XP055481438, ISSN: 1087-0156, DOI: 10.1038/nbt.3803 the whole document	1
A	K. NISHIDA ET AL.: "Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems", SCIENCE, vol. 353, no. 6305, 4 August 2016 (2016-08-04), pages aaf8729-aaf8729, XP055482712, US ISSN: 0036-8075, DOI: 10.1126/science.aaf8729 the whole document	1
A	-& K. NISHIDA ET AL.: "Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems", SCIENCE, vol. 353, no. 6305, 4 August 2016 (2016-08-04), pages aaf8729-aaf8729, XP055482712, US ISSN: 0036-8075, DOI: 10.1126/science.aaf8729 the whole document	1
A	WO 2013/188037 A2 (AGILENT TECHNOLOGIES, INC. [US]) 19 December 2013 (2013-12-19) claim 12 the whole document	49-51
A	YUAN ZONG ET AL.: "Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion", NATURE BIOTECHNOLOGY, vol. 35, no. 5, 27 February 2017 (2017-02-27), pages 438-440, XP055479337, US ISSN: 1087-0156, DOI: 10.1038/nbt.3811 the whole document	62

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/024208

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	-& YUAN ZONG ET AL: "Preci se base edi ting in r i ce, wheat and mai ze wi t h a Cas9-cyti dine deaminase fusion" , NATURE BIOTECHNOLOGY, vol . 35, no. 5, 27 February 2017 (2017-02-27) , pages 438-440, XP055482699 , US ISSN : 1087-0156, D0I : 10. 1038/nbt.3811 the whole document	62
A	KY0UNGMI KIM ET AL. : "Hi ghly effi cient RNA-guided base edi t i ng i n mouse embryos" , NATURE BIOTECHNOLOGY, vol . 35, no. 5, 27 February 2017 (2017-02-27) , pages 435-437, XP055479394, US ISSN : 1087-0156, D0I : 10. 1038/nbt.3816 the whole document	65
A	-& KY0UNGMI KIM ET AL: "Highly effi ci ent RNA-guided base edi t i ng i n mouse embryos" , NATURE BIOTECHNOLOGY, vol . 35, no. 5, 27 February 2017 (2017-02-27) , pages 435-437, XP055482711 , US ISSN : 1087-0156, D0I : 10. 1038/nbt .3816 the whole document	65
A	W0 2016/069910 Al (MASSACHUSETTS EYE & EAR INFI RMARY [US]) 6 May 2016 (2016-05-06) the whole document	127-131
A	KÜRY SÉBASTI EN ET AL. : "De novo di sruption of the proteasome regul atory subuni t PSMD12 causes a syndromi c neurodevel opmental di sorder" , AMERICAN JOURNAL OF HUMAN GENETICS, AMERICAN SOCI ETY OF HUMAN GENETICS, CHICAGO, I L, US, vol . 100, no. 2, 26 January 2017 (2017-01-26) , pages 352-363, XP055482689 , ISSN : 0002-9297, D0I : 10. 1016/J .AJHG.2017.01.003 the whole document	132-136
X, P	W0 2017/070632 A2 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US]) 27 Apri l 2017 (2017-04-27) the whole document	1

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/024208

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P	WO 2018/039438 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US]) 1 March 2018 (2018-03-01) the whole document	1

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/024208

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☑ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-90, 114-116, 137-207 (completely) ; 117, 127-136(partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-90, 114-116, 137-207 (completely) ; 117, 127-136(partially)

A fusion protein comprising: (i) a nucleic acid programmable DNA binding protein (napDNAbp); (ii) a cytidine deaminase domain; and (iii) a uracil glycosylase inhibitor (UGI) domain, wherein the napDNAbp is a CasX, CasY, Cpf1, C2c1, C2c2, C2c3, or Argonaute protein.
A method for editing a nucleobase pair of a double-stranded DNA sequence, the method comprising:
contacting a target region of the double-stranded DNA sequence with a complex comprising a nucleobase editor and a guide nucleic acid, wherein the target region comprises a target nucleobase pair;
inducing strand separation of said target region;
converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase; and
cutting no more than one strand of said target region;
wherein a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase;
wherein the method causes less than 20% indel formation in the double-stranded DNA sequence; and
wherein the nucleobase editor comprises CasX, CasY, Cpf1, C2c1, C2c2, C2c3, or Argonaute.
Products, methods and pharmaceutical compositions related thereto.

---

2. claims: 91-113 (completely) ; 117, 127-136 (partially)

A method for producing a ribonucleoprotein (RNP) complex, the method comprising:
(i) complexing a base editor protein with an RNA in an aqueous solution, thereby forming a complex comprising the base editor and the RNA in the aqueous solution; and
(ii) contacting the complex of (i) with a cationic lipid.
Methods and pharmaceutical compositions related thereto.

---

3. claims: 118-126

A method for purifying a base editor protein, the method comprising:
(i) expressing the base editor protein in a cell, wherein the base editor protein comprises an affinity tag;
(ii) lysing the cell of (i), thereby generating a lysate; and
(iii) subjecting the lysate to affinity chromatography, thereby producing an eluent comprising a purified base editor protein.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2018/024208

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013188037 A2	19-12-2013	US 2013330778 Al	12-12-2013
		WO 2013188037 A2	19-12-2013
W0 2016069910 Al	06-05-2016	EP 3212221 Al	06-09-2017
		EP 3212770 Al	06-09-2017
		US 2017247672 Al	31-08-2017
		US 2017340754 Al	30-11-2017
		W0 2016069910 Al	06-05-2016
		W0 2016069912 Al	06-05-2016
W0 2017070632 A2	27-04-2017	AU 2016342380 Al	10-05-2018
		CA 3002827 Al	27-04-2017
		EP 3365356 A2	29-08-2018
		EP 3365357 A2	29-08-2018
		KR 20180069898 A	25-06-2018
		SG 11201803173V A	30-05-2018
		US 2017121693 Al	04-05-2017
		W0 2017070632 A2	27-04-2017
		W0 2017070633 A2	27-04-2017
W0 2018039438 Al	01-03-2018	NONE	